

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2002, 10:21:53 ; Search time 1822.67 Seconds
(without alignments)
3926.589 Million cell updates/sec

Title: US-09-829-124-1
Perfect score: 342
Sequence: 1 atggactctatcggaacaa.....gcagctggcggaacgcc 342

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1797656 seqs, 10463268293 residues
Total number of hits satisfying chosen parameters: 3595312

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl.*
1: gb_ba.*
2: gb_hgt.*
3: gb_in.*
4: gb_ov.*
5: gb_ov.*
6: gb_pat.*
7: gb_ph.*
8: gb_pl.*
9: gb_pr.*
10: gb_ro.*
11: gb_sts.*
12: gb_sy.*
13: gb_un.*
14: gb_vi.*
15: em_ba.*
16: em_fun.*
17: em_hum.*
18: em_in.*
19: em_mu.*
20: em_mu.*
21: em_or.*
22: em_or.*
23: em_pat.*
24: em_ph.*
25: em_pl.*
26: em_ro.*
27: em_sts.*
28: em_un.*
29: em_un.*
30: em_hgt_hum.*
31: em_hgt_inv.*
32: em_hgt_inv.*
33: em_hgtg_inv.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES
Result No. Score Match Length DB ID Description

1	342	100.0	342	6	AX256351	Sequence
2	49.2	14.4	10096	1	AF026197	Xanthomon
3	48.6	14.2	767	3	EIMMAX	M30933-E.tenella a
4	48	14.0	19304	1	AB045311	AB045311 Xanthomon
5	45	13.2	162869	9	AC090645	AC090645 Homo sapi
6	43.8	12.8	207050	1	AL646063	AL646063 Ralstonia
7	43.2	12.6	190770	2	AC055876	AC055876 Homo sapi
8	43	12.6	178984	2	AC084023	AC084023 Oryza sat
9	42.4	12.4	159971	2	AC091951	AC091951 Homo sapi
10	42.2	12.3	13409	1	AE004466	AE004466 Pseudomon
11	42	12.3	9921	3	AF325500	AF325500 Anthraea
12	42	12.3	142277	2	AC016204	AC016204 Homo sapi
13	41.4	12.1	175296	2	AC106674	AC106674 Rattus no
14	41.2	12.0	204152	2	AC092203	AC092203 Mus muscu
15	40.8	11.9	272545	2	AC090533	AC090533 Mus muscu
16	40.6	11.9	5719	10	MMCATS1	AF051726 Mus muscu
17	40.6	11.9	197265	10	AL603702	AL603702 Mouse DNA
18	40.6	11.9	320096	2	AL603889	AL603889 Mus muscu
19	40.2	11.8	14034	1	AE005694	AE005694 Caulobact
20	40.2	11.8	215046	2	AC011767	AC011767 Homo sapi
21	40.2	11.8	277577	2	AC095395	AC095395 Rattus no
22	40	11.7	72187	2	AC013433	AC013433 Drosophil
23	40	11.7	199508	2	AC098555	AC098555 Rattus no
24	39.8	11.6	10565	1	AE004576	AE004576 Pseudomon
25	39.8	11.6	10729	1	AB012767	AB012767 Pseudomon
26	39.8	11.6	11482	1	AE001876	AE001876 Deinococc
27	39.8	11.6	40510	3	AF067607	AF067607 Caenorhab
28	39.8	11.6	160309	2	AC099443	AC099443 Rattus no
29	39.8	11.6	236962	2	AL606742	AL606742 Mus muscu
30	39.4	11.5	209128	5	AB073377	AB073377 Oryzias l
31	39	11.4	1212	3	PFACSL	M15101 P.cynomolgi
32	39	11.4	110000	2	LMFLCHR34_01	Continuation (2 of
33	39	11.4	190803	9	AL355490	AL355490 Human DNA
34	39	11.4	204050	1	AL646073	AL646073 Ralstonia
35	38.6	11.3	5174	9	HSCOLA43	X80031 Homo sapien
36	38.6	11.3	110000	2	LMFLCHR34_11	Continuation (12 o
37	38.4	11.2	6814	5	AF297658	AF297658 Danio rer
38	38.4	11.2	8009	3	AF226688	AF226688 Bombyx mo
39	38.4	11.2	81053	3	AC005762	AC005762 Drosophil
40	38.4	11.2	89428	2	AC017242	AC017242 Drosophil
41	38.4	11.2	208907	2	AC106604	AC106604 Rattus no
42	38.4	11.2	215050	1	AL646057	AL646057 Ralstonia
43	38.4	11.2	240021	2	AC099436	AC099436 Rattus no
44	38.4	11.2	283698	2	AC099294	AC099294 Rattus no
45	38.4	11.2	302473	3	AE003589	AE003589 Drosophil

ALIGNMENTS

RESULT	1	AX256351	342 bp	DNA	linear	PAT 10-OCT-2001
LOCUS		Sequence 8 from Patent WO0170988.				
DEFINITION		AX256351				
ACCESSION		AX256351.1	GI:16075202			
VERSION						
KEYWORDS						
SOURCE						
ORGANISM						
REFERENCE						
AUTHORS						
TITLE						
JOURNAL						
FEATURES						
source						
BASE COUNT						
ORIGIN						

```
Query Match 100.0%; Score 342; DB 6; Length 342;
Best Local Similarity 100.0%; Pred. No. 1.5e-62;
Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgagctctatcggaacaacttctgaataatcggaacactgcagacgagtcggtcggc 60
Db 1 ATGAGCTCTATCGGAACAACCTTTCGAATATCGCAACCTGCAGACGATGGGATCGG 60

Qy 61 cctcagcaacacagagactcagcagcagtcgcttcgctcggctcggcagcagcagctg 120
Db 61 CCTCAGCAACACAGAGACTCCAGCCAGCAGTCGCTTCGGCTCGGCTCGGACGACGCTG 120

Qy 121 gatcagttgctcgccatgttcacatgatgatgctcaacagacagccagcagcagtcga 180
Db 121 GATCAGTTGCTCGCCATGTTTCATCATGATGATGCTGCAACAGACCCAGGCGCGATGCA 180

Qy 181 aatcaggagtggtggaacacacagcagcagtcgcttcgctcggcagcagcagtcggtg 240
Db 181 AATCAGGAGTGTGGCAACGAACAACCGCAGACGCTCAACAGGAAGCCTGAGTCGCTTG 240

Qy 241 accagatgctgatgacagtcgctgatgacagtcgctgatgacagcagcagcagcagcag 300
Db 241 ACCAGATGCTGATGACAGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 300

Qy 301 ggcgggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 342
Db 301 GCGGGTGGCGGTGCGGTGCAACAGCAGCCTGGCGGCAACGCC 342

RESULT 2
AF026197 10096 bp DNA linear BCT 30-MAR-2000
LOCUS Xanthomonas oryzae pv. oryzae hrp gene cluster, partial sequence.
DEFINITION AF026197 AF232057
ACCESSION AF026197.2 GI:7350908
VERSION
KEYWORDS
SOURCE Xanthomonas oryzae pv. oryzae.
ORGANISM Xanthomonas oryzae pv. oryzae
Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
Xanthomonas.
REFERENCE 1 (bases 1 to 10096)
AUTHORS Zhu, W., Magbanua, M.M. and White, F.F.
TITLE Identification of two novel hrp-associated genes in the hrp gene
cluster of xanthomonas oryzae pv. oryzae
J. Bacteriol. 182 (7), 1844-1853 (2000)
MEDLINE 20179797
PUBMED 10714988
REFERENCE 2 (bases 891 to 2420)
AUTHORS Zhu, W. and White, F.F.
TITLE HrpXo is a regulator of hrp gene expression in Xanthomonas oryzae
pv. oryzae
Unpublished
REFERENCE 3 (bases 891 to 2420)
AUTHORS Zhu, W. and White, F.F.
TITLE Direct Submission
Submitted (22-SEP-1997) Plant Pathology, Kansas State University,
Throckmorton Hall, Manhattan, KS 66506, USA
REFERENCE 4 (bases 1 to 10096)
AUTHORS White, F.F., Zhu, W. and Magbanua, M.M.
TITLE Direct Submission
Submitted (08-FEB-2000) Plant Pathology, Kansas State University,
4024 Throckmorton Hall, Manhattan, KS 66506, USA
REMARK Sequence update by submitter
COMMENT On Mar 30, 2000 this sequence version replaced gi:4003499.
FEATURES
source
1..10096
/organism="Xanthomonas oryzae pv. oryzae"
/strain="PX086"
/db_xref="taxon:64187"
/feature="pathovar: oryzae"
misc_feature
1..6
/feature="adjoins with sequence presented in GenBank"
```

```
Accession Number AF232058"
Complement(129..908)
/gene="hpa2"
complement(129..617)
/gene="hpa2"
/note="lysosome related protein; hrp-associated protein"
/codon_start=1
/transl_table=11
/evidence="not experimental"
/product="Hpa2 precursor"
/protein_id="AAF61278.1"
/db_xref="GI:7350910"
/translation="MLFAALACAAAFARACDCEEAAGYOHVNPWLRLATAWQESRGR
ADAIHRNNNGTVDYGRMQINSIHLRLFGGISKALMQPCVSYVAARLREMTNKY
GNTAAVAGAYHSETPGERDKYAHAIHSILRLGLGVNVRASDPAGMRGAAMRMRHLHQ
VA"
complement(570..617)
/gene="hpa2"
complement(884..908)
/gene="hpa2"
/note="variant plant-inducible promoter, Pip-1"
976..1000
/gene="hpa1"
/note="variant plant-inducible promoter, Pip-2"
976..1567
/gene="hpa1"
1136..1567
/gene="hpa1"
/note="Orf432; HrpXo; similar to Xanthomonas oryzae pv.
oryzae PopA; hrp associated protein; glycine-rich protein
encoded by hrp region"
/codon_start=1
/transl_table=11
/product="Hpa1"
/protein_id="AAC95121.1"
/db_xref="GI:4003500"
/translation="MNSLNTQFGSTNLQVPSQDTTFGSNGGNOGISEKOLDQLL
COLISALQSSKNAEKGKGGGNGGQSQGQNGPSPFTOMLHIVEILQQA
QNGGAGGGGFGGFGGFGGSGDLGLTNLSSGSASMQ"
complement(2306..4123)
/gene="hrpA"
complement(2306..4123)
/gene="hrpA"
/note="Hrcc related"
/codon_start=1
/transl_table=11
/product="Hrpa"
/protein_id="AAC95122.2"
/db_xref="GI:7350909"
/translation="MAPACTTTHRQAPLFAVILLSLFLSPQADAQVPHSRTFK
YVADNKDLKVLRLDLSASQSIATWISPEVTGLSGKFETSPQKFLDLDLAATYGFVYVY
DGAMLRWGANESKATSLGTSKSLRDLRLDSDRPVRYDEAAHVAVVSGP
PGYVDTVSAIAKQVEQARORDATEVQVQALHYAQRADHTTRIGQDVQIPGASLLR
SIYGARGASVAFIAAFGANFGRVQPIGGSSNTFGNAOCCGGGASGTLGLSPFNGG
ASPDPMFVSPPLPGSGAAGSPASVMPSELKSRDESNPIDAGGAELASDAPVIEA
DPTNAILIRDRPERMQSYGTLIQDLNRPKLIQDITIEIRDMGQMDLGYDWRFS
QHTDITQDGGGGLGFGNGLSAAATDGTATPVGTLTAVLGDAGRYLMTRVSALETT
NKAKTVSSPOVATLDNVEAMDKQAFVRSVGSYASADLYNLGASVSLRVLPSPVPGS
PNGOMLRVRIEDGOLGSNVVDGIPVITSSEIKTQAFVNEGQSLIAGYAYDADETDL
NAVGLSKIPLLGNLFKIRQSGSRMQRLLLTTPHVSF"
complement(4206..5036)
/gene="hrpB"
complement(4206..9926)
/gene="hrpB operon"
complement(4206..5036)
/gene="hrpB"
/codon_start=1
/transl_table=11
/product="HrpB"
/protein_id="AAF61279.1"
/db_xref="GI:7350911"
/translation="MNDVTDALLSSQGVLLTLLALCGRVFVFWFVLPATQDLSL
PGIARNGVIYVLLSFIAYGQPDALAKITQTVGLGVGWEKFAFILLIGFAAFWIA
```



```

/gene="hrpB5"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="HrpB5"
/protein_id="BAB07853.1"
/db_xref="GI:9711587"
/translacion="MRVLRSTPDIGDGVWPREALASVLALDAALAAVEVHARCEQA
LSQAARQTLIEAQQOAEAILHDAROKAERSARLGYAAGLRQLDWNESSGLRHAF
AAETAARAFERLAEIVARTCEHVILGHDPALYARAQALEGALDENKALRVSVHPE
ALDAARAFDAAYKAGWTLOVELCGDLAVGACVCEWDTGVFTDLRDLQRLSRRV
IRRVLAQPEADVG"
complement(6617. 7246)
/gene="hrpB4"
complement(6617. 7246)
/gene="hrpB4"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="HrpB4"
/protein_id="BAB07854.1"
/db_xref="GI:9711585"
/translacion="MDNTRIDNARCVRQRTDALAQLAESLDAAARVAHWLGPVAASGW
AAPVYRMRYIQWGSSELSPALGVHANQALALMPARLANLIARALFSRGLAVR
RCIERLGLWLEQCVPAVLEHVRONASGLSVPLPRDADQAAWGDGWRMRQADGV
WSNPFVAKLLSLPLGAAQVATVAADGASAAFTLPTLMPAESCVSG"
complement(7254. 8015)
/gene="hrpB3"
complement(7254. 8015)
/gene="hrpB3"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="HrpB3"
/protein_id="BAB07855.1"
/db_xref="GI:9711586"
/translacion="MRALRYLVLLALLSACSQQLYSGLTENDANDMLVLLHAGVD
ASKVTPDDGTWVAPNAPHQDVYSLEALRAHGLPHERHANLGMFKDGLISTPTER
VRIYGVQSOQLSTNIDGIVSADVEIPLPNNDPLATSVKPSAAVFIKFRVGSIDIT
SLVPNTKTYMHVSVEGLTIVENVSVTLVPGCAESDAQLTASAPRPPSPWPVWVCVTL
CLAGAALYWPNPQAGRWGGLQRLRLITTKRAG"
complement(8017. 8409)
/gene="hrpB2"
complement(8017. 8409)
/gene="hrpB2"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="HrpB2"
/protein_id="BAB07856.1"
/db_xref="GI:9711587"
/translacion="WLIIPVQVIACTSAATQALSPVAPNQAALVNRFOALMQSSSP
LPPMQRVGSPPSMRSRVVDQNDVETIAEHIDAFSMQAPTMLQEMAAQOIKLMHEL
TWGFFNLVSVGVAQSGKNAVQTLVKNQ"
complement(8443. 8898)
/gene="hrpB1"
complement(8443. 8898)
/gene="hrpB1"
/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="HrpB1"
/protein_id="BAB07857.1"
/db_xref="GI:9711588"
/translacion="MEKIQCGSVVSGLIETIVGLTHEKIQDAALAAVRLRPEL
KALDFTDAMISIKRNYVDGLRELEGDAGSKPLCRALYACCLFAMGDPSWHGVAD
GLIEDADADAVALKALSGRTTPTTAPADVPVSSAPNEVPSQYLRA"
9113. 10192
/gene="hrpC"
9113. 17604
/feature="hrp gene cluster 2"
9113. 10192
/gene="hrpC"

```

```

/codon_start=1
/transl_table=11
/evidence-not_experimental
/product="HrcU"
/protein_id="BAB07858.1"
/db_xref="GI:9711589"
/translacion="MSEKTEKPEKLRDKDEVPVSDVTAAYVLCALLVMKS
AGDFYDHRALTRIGDFSENTRDATALRALAHIGLOGLLMLPFLAACLVAGLVG
GAFTGPNASLKPSPKPSLNPANGVKKLSRLSLINLKLIIKAILIGVMMVGIR
TLMPTIGLAYETLIDISQIAWRTLSMLFALGVLLILVGAADNSVQHWLFTRDKRMS
Query Match 14.0%; Score 48; DB 1; Length 19304;
Best Local Similarity 54.5%; Pred No. 1.6;
Matches 96; Conservative 0; Mismatches 80; Indels 0; Gaps 0;
QY 161 agagccagggcagcgatgcaaatcaggagtggtggcaacgaaacccgagaaaggtcaac 220
||| ||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 410 AGGTCAGGGTGGCGATAATGCGGTGGCAGGCGGCAATTCGCAGCAGCGCGGCGAC 469
||| ||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 221 aggaagggcgtgctcgttgacgcagatgctgagtcagatcgtgagtcagtcagcaga 280
||| ||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 470 AGAATGGCCCTCGCCATTCACCCAGATGCTGATCATATCTCGCGAGATTCCTCCAG 529
||| ||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 281 accagggcgcgccgcagtcggtgctggttcggttcggttcggttcggttcggttcggt 336
||| ||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 530 CGCAGATGGTGGTGGTCTGCTGGTGGCGGGGTTCCGGCGGGGTTCCGGCGGTGAC 585
||| ||||| ||||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
AC090645 162869 bp DNA linear PRI 06-MAR-2001
LOCUS Homo sapiens chromosome 3 clone RP11-556G18 map 3p, complete
sequence.
AC090645
VERSION AC090645.1 GI:13236634
KEYWORDS HTG.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
1 (bases 1 to 162869)
AUTHORS
Luo,C., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C.,
Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
Liu,Y., Li,W., Li,Y., Luo,J., Luo,J., Liu,Y., Qi,Q., Qi,X., Song,L.,
Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Fan,C., Yang,X.,
Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
Yu,J. and Yang,H.
Chromosome 3p genomic sequence
Unpublished
2 (bases 1 to 162869)
AUTHORS
Luo,C., Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J.,
Ding,H., Dong,W., Fan,H., Feng,X., Gong,J., Guan,Q., Gu,X., Guo,D.,
Guo,Z., He,L., Hu,S., Huang,F., Jin,Y., Kang,N., Li,C., Li,C.,
Li,F., Li,G., Li,J., Li,L., Li,S., Li,T., Liu,Y., Liu,N., Liu,B.,
Liu,Y., Li,W., Li,Y., Luo,J., Luo,J., Liu,Y., Qi,Q., Qi,X., Song,L.,
Song,S., Sun,M., Sun,W., Sun,Y., Tan,X., Tao,R., Wang,H., Wang,J.,
Wang,J., Wang,J., Wang,L., Wang,L., Wang,R., Wang,X., Wang,X.,
Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y., Fan,C., Yang,X.,
Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H., Zhang,L., Zhang,M.,
Zhang,X., Zhang,X., Zhang,Y., Zhang,Y., Zhang,Z., Zhu,B., Zhu,N.,
Yu,J. and Yang,H.
Direct Submission
Submitted (06-MAR-2001) Human Genomic Center, Institute of
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing
100101, P.R.China
-----Genome Center
Center:Beijing Center
Center code:Beijing

```


estimation.

```

/transl_table=11
/product="hynothetical protein"

```

estimation.

Quiles, M., Ren, Y., Rives, M., Rojas, A., Rojibokan, I., Rolfe, M.,
Ruiz, S., Savary, G., Scherer, S., Scott, G., Shen, H., Shoostari, N.,
Sisson, I., Sodergren, E., Sonaike, T., Sparks, A., Stanley, H.,
Stang, H., Sutton, A., Svatek, A., Tabor, P., Tamerisa, A., Tamerisa, K.,
Tang, H., Tansey, J., Taylor, C., Taylor, T., Telford, B., Thomas, N.,
Thomas, S., Usmani, K., Vasquez, L., Vera, V., Villalon, D., Vinson, R.,
Wall, R., Wang, S., Ward-Moore, S., Warren, R., Washington, C.,
Watlington, S., Williams, G., Williamson, A., Wleczyk, R., Woodson, S.,
Worley, K., Wu, C., Wu, Y., Wu, Y.F., Zhou, J., Zorrilla, S., Nelson, D.,
Weinstock, G. and Gibbs, R.
Direct Submission

TITLE

JOURNAL
REFERENCE
AUTHORS

2 (bases 1 to 175296)

Worley, K.C.

TITLE

Submitted (12-JAN-2002) Human Genome Sequencing Center, Department
of Molecular and Human Genetics, Baylor College of Medicine, One
Baylor Plaza, Houston, TX 77030, USA

COMMENT

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: G10J

Center clone name: CH230-108D18

----- Summary Statistics

Assembly program: Phrap; version 0.990329first call to

findPhrapList

Consensus quality: 148335 bases at least Q40
Consensus quality: 155812 bases at least Q30
Consensus quality: 162105 bases at least Q20
Estimated insert size: 152728; sum-of-contigs estimation
Quality coverage: 0x in Q20 bases; agarose-fp estimation
Quality coverage: 2.4x in Q20 bases; sum-of-contigs estimation

----- NOTE: Estimated insert size may differ from sequence length
(see http://www.hgsc.bcm.tmc.edu/docs/genbank_draft_data.html).
NOTE: This is a 'working draft' sequence. It currently
consists of 56 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 14942: contig of 14942 bp in length
14943 15042: gap of unknown length
15043 26820: contig of 11778 bp in length
26821 26920: gap of unknown length
26921 31321: contig of 4401 bp in length
31322 31421: gap of unknown length
31422 37389: contig of 5968 bp in length
37390 37489: gap of unknown length
37490 42498: contig of 5009 bp in length
42499 42598: gap of unknown length
42599 46308: contig of 3710 bp in length
46309 46408: gap of unknown length
46409 52395: contig of 5987 bp in length
52396 52495: gap of unknown length
52496 59139: contig of 6644 bp in length
59140 59239: gap of unknown length
59240 64327: contig of 5088 bp in length
64328 64427: gap of unknown length
64428 68633: contig of 4206 bp in length
68634 68733: gap of unknown length
68734 73328: contig of 4595 bp in length
73329 73428: gap of unknown length
73429 77037: contig of 3609 bp in length
77038 77137: gap of unknown length
77138 81483: contig of 4346 bp in length
81484 81584: gap of unknown length
81584 85544: contig of 3961 bp in length

85545 85644: gap of unknown length
85645 89737: contig of 4093 bp in length
89738 89837: gap of unknown length
89838 94559: contig of 4722 bp in length
94559 94659: gap of unknown length
94660 96754: contig of 2095 bp in length
96755 96854: gap of unknown length
96855 98841: contig of 2987 bp in length
98842 99941: gap of unknown length
99942 103396: contig of 3455 bp in length
103397 103496: gap of unknown length
103497 106178: contig of 2682 bp in length
106179 106278: gap of unknown length
106279 109015: contig of 2737 bp in length
109016 109115: gap of unknown length
109116 111536: contig of 2421 bp in length
111537 111636: gap of unknown length
111637 114747: contig of 3111 bp in length
114748 114847: gap of unknown length
114848 117377: contig of 2530 bp in length
117378 117477: gap of unknown length
117478 120113: contig of 2636 bp in length
120114 122860: gap of unknown length
122861 122960: gap of unknown length
122961 125473: contig of 2513 bp in length
125474 125573: gap of unknown length
125574 128271: contig of 2898 bp in length
128272 128371: gap of unknown length
128372 131331: contig of 2960 bp in length
131332 131431: gap of unknown length
131432 133677: contig of 2246 bp in length
133678 133777: gap of unknown length
133778 135703: contig of 1926 bp in length
135704 135803: gap of unknown length
135804 137886: contig of 1983 bp in length
137887 137886: gap of unknown length
137887 140617: contig of 2731 bp in length
140618 140717: gap of unknown length
140718 141934: contig of 1217 bp in length
141935 142034: gap of unknown length
142035 144079: contig of 2045 bp in length
144080 144179: gap of unknown length
144180 146430: contig of 2251 bp in length
146431 146530: gap of unknown length
146531 148261: contig of 1731 bp in length
148262 148361: gap of unknown length
148362 150017: contig of 1656 bp in length
150018 150117: gap of unknown length
150118 151592: contig of 1475 bp in length
151593 151692: gap of unknown length
151693 153207: contig of 1515 bp in length
153208 153307: gap of unknown length
153308 154357: contig of 1050 bp in length
154358 154457: gap of unknown length
154458 156066: contig of 1609 bp in length
156067 156166: gap of unknown length
156167 157546: contig of 1379 bp in length
157546 157645: gap of unknown length
157646 159211: contig of 1566 bp in length
159212 159311: gap of unknown length
159312 160381: contig of 1070 bp in length
160382 160481: gap of unknown length
160482 161912: contig of 1431 bp in length
161913 162012: gap of unknown length
162013 163616: contig of 1604 bp in length
163617 163716: gap of unknown length
163717 164732: contig of 1016 bp in length
164733 164832: gap of unknown length
164833 165958: contig of 1126 bp in length
165959 166058: gap of unknown length
166059 167610: contig of 1552 bp in length
167611 167710: gap of unknown length

```

* 167711 168803: contig of 1093 bp in length
* 168804 168903: gap of unknown length
* 168904 169071: contig of 168 bp in length
* 169072 170071: gap of unknown length
* 170072 171105: contig of 1034 bp in length
* 171106 172205: gap of unknown length
* 172206 172289: contig of 1084 bp in length
* 172290 172389: gap of unknown length
* 172390 173513: contig of 1124 bp in length
* 173514 173613: gap of unknown length
* 173614 175296: contig of 1683 bp in length.
FEATURES             Location/Qualifiers
  source              1..175296
Query Match
Best Local Similarity 12.1%; Score 41.4; DB 2; Length 175296;
Matches 146; Conservative 0; Mismatches 156; Indels 1; Gaps 1;
Qy 38 acctgacagcagtgagcctggcctcagcaacagagagactccagccagcagctgcctt 97
Db 141423 AGCAGCAGGAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 141364
Qy 98 cggctggtccgagcagcagtgatcagttgctgcctatgttcctcatgtatgtatgc 157
Db 141363 AGCAGCAGGAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 141304
Qy 158 aacagagccagcagcagcagcagcagcagcagcagcagcagcagcagcagcagc 217
Db 141303 AGCAGGAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 141244
Qy 218 aacaggaagcctgagtcgctgacagcagcagcagcagcagcagcagcagcagc 277
Db 141243 AGGAGCAGGAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 141184
Qy 278 agaaccagg-gcggcgccgagcagcagcagcagcagcagcagcagcagcagcagc 336
Db 141183 AGGAGCAGGAGCAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 141124
Qy 337 aac 339
Db 141123 AGC 141121

```

```

RESULT 14
AC092203
LOCUS
DEFINITION
Mus musculus clone rp23-422n18 strain C57BL/6J, WORKING DRAFT
SEQUENCE, 11 unordered pieces.
AC092203
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
house mouse.
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
Jia H., Zhang P., Lin S. and Roe, B.A.
TITLE
Mus musculus BAC Clone rp23-422n18
JOURNAL
Unpublished
REFERENCE
2 (bases 1 to 204152)
Jia H., Zhang P., Lin S. and Roe, B.A.
Direct Submission
Submitted (28-JUN-2001) Department of Chemistry And Biochemistry,
The University of Oklahoma, 620 Parrington Oval, Room 208, Norman,
OK 73019, USA
COMMENT
On Jan 29, 2002 this sequence version replaced gi:18139426.
----- Genome Center
Center: Department of Chemistry And Biochemistry
The University of Oklahoma
Center code:UOKNOR
-----

```

* NOTE: This is a 'working draft' sequence. It currently consists of 11 contigs. The true order of the pieces

* is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. * This record will be updated with the finished sequence * as soon as it is available and the accession number will * be preserved.

```

* 1
* 2023: contig of 2023 bp in length
* 2024 2123: gap of unknown length
* 2124 4788: contig of 2665 bp in length
* 4789 4888: gap of unknown length
* 4889 7174: contig of 2286 bp in length
* 7175 7274: gap of unknown length
* 7275 10351: contig of 3077 bp in length
* 10352 10451: gap of unknown length
* 10452 13314: contig of 2863 bp in length
* 13315 13414: gap of unknown length
* 13415 16003: contig of 2589 bp in length
* 16004 16103: gap of unknown length
* 16104 18443: contig of 2340 bp in length
* 18444 18543: gap of unknown length
* 18544 24743: contig of 6200 bp in length
* 24744 24843: gap of unknown length
* 24844 62943: contig of 38100 bp in length
* 62944 63043: gap of unknown length
* 63044 97892: contig of 34849 bp in length
* 97893 97992: gap of unknown length
* 97993 204152: contig of 106160 bp in length.
FEATURES             Location/Qualifiers
  source              1..204152

```

```

/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="rp23-422n18"
/clone_lib="RPCI mouse BAC library 23"

```

```

BASE COUNT 60859 a 43428 c 41581 g 57282 t 1002 others
ORIGIN

```

```

Query Match 12.0%; Score 41.2; DB 2; Length 204152;
Best Local Similarity 49.1%; Pred. No. 36;
Matches 109; Conservative 0; Mismatches 113; Indels 0; Gaps 0;
Qy 64 cagcaacacagaggactccagcagcagcagcagcagcagcagcagcagcagcagc 123
Db 106646 CACCAGCAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 106705
Qy 124 cagttgctccatgttcctcatgtatgtctgcaacagagccagcagcagcagcagc 183
Db 106706 CAGCAGCAGGAGCAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 106765
Qy 184 caggagtgtgcaacgaaacccgcaacggttcaacaggaagcctgagtcggttgcg 243
Db 106766 CAGCAGGAGCAGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 106825
Qy 244 cagatgctgtatgcagatcgtgtatgcagcagcagcagcagcagcagcagcagc 285
Db 106826 CAGCAGCAGGAGCAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 106867

```

```

RESULT 15
AC090533/c
LOCUS
DEFINITION
Mus musculus clone RP23-33H13 strain C57BL6/J, WORKING DRAFT
SEQUENCE, 45 unordered pieces.
AC090533
AC090533.4 GI:17976439
VERSION
HTG; HTGS_PHASE1; HTGS_DRAFT.
KEYWORDS
house mouse.
SOURCE
Mus musculus
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
1 (bases 1 to 272545)
Li, L., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M.,

```

Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
 Perera, A., Shim, C., Thomas, E. and Kucheriapatti, R.
 High Throughput Mouse Sequencing
 2 (bases 1 to 272545)
 Li, L., Montgomery, K.T., Grills, G., Chiu, D., Decker, J., Fusina, M.,
 Goltz, J., Halder, A., Hall, L., Han, J., Ioshikhes, I.P., Lee, E.,
 Perera, A., Shim, C., Thomas, E. and Kucheriapatti, R.
 Direct Submission
 Submitted (02-MAR-2001) Department of Molecular Genetics, Albert
 Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
 Bronx, NY 10461, USA
 On Dec 21, 2001 this sequence version replaced gi:13310871.
 -----Genome Center
 Center: Harvard Partners Genome Center
 Center Code: HPGC
 Web site:
 http://wchanning.bwh.harvard.edu:9088/hpcgg/jsp/hpcgg/Sequence/mous
 e.html
 Contact: gntm@capecod.bwh.harvard.edu
 -----Summary Statistics

Center project name: ADU
 Sequencing vector: pUC18; L08752
 Chemistry: Dye-terminator Big Dye; 100%
 *Consensus quality: 257441 at least Q20
 *Consensus quality: 253718 at least Q30
 *Consensus quality: 247095 at least Q40
 Estimated insert size: agarose-FP - N/A
 **Estimated insert size: 271665 - sum-of-contigs
 Quality coverage: agarose-FP - N/A
 Quality coverage: 6.8 x in Q20 bases; sum-of-contigs estimation

* NOTE: This is a 'working draft' sequence. It currently
 * consists of 45 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

1 30267: contig of 30267 bp in length
 * 30268: gap of unknown length
 * 30288: contig of 28113 bp in length
 * 58400: gap of unknown length
 * 58421: contig of 30444 bp in length
 * 88865: gap of unknown length
 * 88885: contig of 21005 bp in length
 * 109890: gap of unknown length
 * 109910: contig of 15224 bp in length
 * 125133: gap of unknown length
 * 125134: contig of 14396 bp in length
 * 139554: gap of unknown length
 * 139559: gap of unknown length
 * 139570: contig of 13841 bp in length
 * 153411: gap of unknown length
 * 153431: contig of 9137 bp in length
 * 162568: gap of unknown length
 * 162587: gap of unknown length
 * 171045: contig of 8457 bp in length
 * 171065: gap of unknown length
 * 179305: contig of 8240 bp in length
 * 179325: gap of unknown length
 * 187059: contig of 7735 bp in length
 * 187060: gap of unknown length
 * 187080: contig of 10519 bp in length
 * 197599: gap of unknown length
 * 197618: gap of unknown length
 * 205966: contig of 8348 bp in length
 * 205967: gap of unknown length
 * 205987: contig of 6463 bp in length
 * 212450: gap of unknown length
 * 212470: contig of 5283 bp in length
 * 217753: gap of unknown length
 * 217773: contig of 6171 bp in length
 * 223944: gap of unknown length
 * 223963: contig of 6399 bp in length
 * 230362: contig of 5869 bp in length

230363: gap of unknown length
 * 230383: contig of 5869 bp in length
 * 236251: contig of 5869 bp in length
 * 236252: gap of unknown length
 * 236271: contig of 4301 bp in length
 * 240572: gap of unknown length
 * 240573: gap of unknown length
 * 240593: contig of 2197 bp in length
 * 242789: contig of 1886 bp in length
 * 242809: gap of unknown length
 * 242810: contig of 1886 bp in length
 * 244696: gap of unknown length
 * 244715: gap of unknown length
 * 244716: contig of 2254 bp in length
 * 246969: gap of unknown length
 * 246970: contig of 881 bp in length
 * 246990: gap of unknown length
 * 247871: contig of 303 bp in length
 * 247890: gap of unknown length
 * 247891: contig of 303 bp in length
 * 248194: gap of unknown length
 * 248213: contig of 1909 bp in length
 * 248214: gap of unknown length
 * 250123: contig of 1225 bp in length
 * 250142: gap of unknown length
 * 250143: contig of 1225 bp in length
 * 251368: gap of unknown length
 * 251388: contig of 1362 bp in length
 * 252749: gap of unknown length
 * 252750: contig of 871 bp in length
 * 252750: gap of unknown length
 * 252770: contig of 871 bp in length
 * 253641: gap of unknown length
 * 253660: gap of unknown length
 * 253661: contig of 1372 bp in length
 * 255032: gap of unknown length
 * 255033: contig of 2281 bp in length
 * 255053: gap of unknown length
 * 257333: gap of unknown length
 * 257334: contig of 2056 bp in length
 * 257354: gap of unknown length
 * 259410: contig of 1030 bp in length
 * 259429: gap of unknown length
 * 259430: contig of 1030 bp in length
 * 260460: gap of unknown length
 * 260479: gap of unknown length
 * 260829: contig of 350 bp in length
 * 260830: gap of unknown length
 * 260849: contig of 812 bp in length
 * 261661: gap of unknown length
 * 261681: contig of 851 bp in length
 * 262532: gap of unknown length
 * 262533: contig of 799 bp in length
 * 262553: gap of unknown length
 * 263351: contig of 799 bp in length
 * 263352: gap of unknown length
 * 263372: contig of 846 bp in length
 * 264217: gap of unknown length
 * 264237: contig of 968 bp in length
 * 264238: gap of unknown length
 * 265205: contig of 1045 bp in length
 * 265225: gap of unknown length
 * 265226: contig of 459 bp in length
 * 266271: gap of unknown length
 * 266290: gap of unknown length
 * 266749: gap of unknown length
 * 266750: contig of 949 bp in length
 * 266770: gap of unknown length
 * 267719: contig of 838 bp in length
 * 267739: gap of unknown length
 * 268576: gap of unknown length
 * 268577: contig of 1414 bp in length
 * 268597: gap of unknown length
 * 270011: contig of 1414 bp in length
 * 270031: gap of unknown length
 * 271654: contig of 1624 bp in length
 * 271655: gap of unknown length
 * 271675: contig of 871 bp in length.
 * 272545: Location/Qualifiers
 1. 272545
 /organism="Mus musculus"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="RP23-33H13"
 /sex="male"
 1. 30267
 /note="assembly_name:Contig108"
 30288. 58400
 /note="assembly_name:Contig107"
 clone_end:SP6
 vector_side:left
 58421. 88864
 /note="assembly_name:Contig106"
 88885. 109889
 /note="assembly_name:Contig105"

FEATURES
 source

misc_feature
 misc_feature
 misc_feature
 misc_feature

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2002, 10:22:48 ; Search time 177.26 Seconds
(without alignments)
3312.562 Million cell updates/sec

Title: US-09-829-124-1

Perfect score: 342

Sequence: 1 atggactctatcggaacaa.....gcagctggcggaacgcc 342

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1736436 seqs, 858457221 residues

Total number of hits satisfying chosen parameters: 3472872

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_032802.*
1: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1980.DAT.*
2: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1981.DAT.*
3: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1982.DAT.*
4: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1983.DAT.*
5: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1984.DAT.*
6: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1985.DAT.*
7: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1986.DAT.*
8: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1987.DAT.*
9: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1988.DAT.*
10: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1989.DAT.*
11: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1990.DAT.*
12: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1991.DAT.*
13: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1992.DAT.*
14: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1993.DAT.*
15: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1994.DAT.*
16: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1995.DAT.*
17: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1996.DAT.*
18: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1997.DAT.*
19: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1998.DAT.*
20: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA1999.DAT.*
21: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2000.DAT.*
22: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT.*
23: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT.*
24: /SIDS1/gcgdata/geneseq/geneseq-emb1/NA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	342	100.0	342	22	AAD18127
2	38.8	11.3	2004	18	AT85356
3	38.4	11.2	1925	20	AA90924
4	38.4	11.2	3840	23	ABL23322
5	38	11.1	3121	21	AA51260
6	37.8	11.1	1581	22	AAF61055
7	37.8	11.1	4335	23	AA579854
8	37.8	11.1	5120	22	AA84677
9	37.6	11.0	1035	18	AA749316

10	37.6	11.0	1035	19	AAV54609
11	37.6	11.0	1035	19	AAV39975
12	37.6	11.0	1035	19	AAV36430
13	37.6	11.0	1035	20	AAV83991
14	37.6	11.0	1035	21	AA00674
15	37.6	11.0	1035	21	AA14944
16	37.6	11.0	1035	22	AA12808
17	37.4	10.9	1650	23	ABL23323
18	37.4	10.9	2444	23	ABL05117
19	37.4	10.9	4444	23	ABL05116
20	37.2	10.9	447	23	ABL27665
21	37.2	10.9	506	21	AA42475
22	37.2	10.9	574	22	AA101496
23	37.2	10.9	2338	12	AAQ14183
24	37.2	10.9	2338	19	AAV23249
25	37.2	10.9	2338	21	AA238195
26	37.2	10.9	2744	16	AAQ98470
27	37.2	10.9	3477	23	ABL27664
28	37.2	10.9	4203	22	AA05260
29	37	10.8	900	23	AA51554
30	37	10.8	6463	21	AAZ45239
31	36.8	10.8	477	22	AA101459
32	36.6	10.7	2177	21	AAZ57853
33	36.6	10.7	6155	23	ABL09166
34	36.6	10.7	6955	22	AAK72397
35	36.6	10.7	13010	22	AAK72395
36	36.4	10.6	317	21	AA021116
37	36.4	10.6	821	22	AAH05110
38	36.4	10.6	4105	22	AA161196
39	36.4	10.6	4396	22	AAH17588
40	36.4	10.6	4428	22	AA159410
41	36.4	10.6	4403765	22	AA199683
42	36.2	10.6	396	22	AA181388
43	36.2	10.6	423	22	AA164895
44	36	10.5	1126	19	AAV29159
45	35.8	10.5	4262	21	AA064577

ALIGNMENTS

RESULT 1

AAD18127
ID AAD18127 standard; cDNA; 342 BP.

XX AAD18127;

XX 18-DEC-2001 (first entry)

XX Xanthomonas campestris hypersensitive response elicitor protein cDNA.

XX Plant pathogen; hypersensitive response elicitor; HRE; stress resistance; disease resistance; Erwinia amylovora hypersensitive response elicitor; harpin; signal transduction; growth enhancement; insect control; virucide; fungicide; antibacterial; ss.

XX Xanthomonas campestris.

XX key Location/Qualifiers
CDS 1..342

XX /tag= a
XX /product= "Xanthomonas campestris hypersensitive response elicitor protein"
XX /note= "CDS does not include stop codon"
XX /partial

XX WO200170988-A2.

XX 27-SEP-2001.

XX 19-MAR-2001; 2001WO-US08728.

XX 23-MAR-2000; 2000US-191649P.

Pseudomonas solana
Pseudomonas solana
Hypersensitive res
DNA encoding a hyp
Pseudomonas solana
DNA encoding a hyp
Pseudomonas solana
Drosophila melanog
Drosophila melanog
Drosophila melanog
Human secreted exp
Human reproductive
N. clavipes draglin
N. clavipes spider
MISPI-containing p
Drosophila melanog
Human reproductive
Pseudomonas aerugi
DNA encoding angio
Human reproductive
Protein regulating
Drosophila melanog
Human immune/haema
Human immune/haema
Human secreted pro
Human cDNA clone (
Human polynucleoti
Human cDNA sequenc
Human polynucleoti
Mycobacterium tube
Human polynucleoti
Xanthomonas hrfl p
Nucleotide sequenc
E. tenella CGMP de

PR 01-DEC-2000; 2000US-250710P.
 XX (EDEN-) EDEN BIOSCIENCE CORP.
 XX
 XX Song X, Fan H, Wei Z;
 XX WPI: 2001-590177/66.
 DR P-PSDB; AAE10807.
 XX
 XX New plant pathogen hypersensitive response elicitor-receptor protein
 PT isolated from plants, which upon silencing is used to study plant
 PT signal transduction pathways leading to disease resistance and growth
 PT enhancement
 XX
 XX Disclosure: Page 19; 78pp; English.
 XX
 XX The present invention relates to an isolated protein which serves as a
 CC receptor in plants for plant pathogen hypersensitive response elicitors
 CC (HRE). The plants are made resistant to infection by viruses, bacteria
 CC and fungi and are imparted with resistance against environmental stress
 CC and insects through HRE treatment. The protein is useful for
 CC understanding the harpin (Erwinia amylovora hypersensitive response
 CC elicitor) induced signal transduction pathway in plants. The protein is
 CC useful for studying the downstream components of signal transduction
 CC pathway in plants which eventually leads to disease resistance, growth
 CC enhancement, insect control and stress resistance. Imparting disease
 CC resistance to plants through HRE treatment has the potential to treat
 CC previously untreatable diseases, treating diseases systemically which
 CC might not be treated separately due to cost and avoids the use of
 CC infectious agents or environmentally harmful materials. By HRE treatment
 CC enhanced plant growth is achieved which includes greater yield,
 CC increased quantity of seeds produced, increased percentage of seeds
 CC germinated, increased plant size, greater biomass, more and bigger
 CC fruits, etc. which results in economic benefit to cultivators. Greater
 CC yield, increased size and enhanced biomass allow greater revenue
 CC generation from the given plot of plant. The present sequence is
 CC *Xanthomonas campestris* pv. *pelargonii* hypersensitive response elicitor
 CC protein cDNA.
 XX
 XX Sequence 342 BP; 83 A; 94 C; 109 G; 56 T; 0 other;

Query Match 100.0%; Score 342; DB 22; Length 342;
 Best Local Similarity 100.0%; Pred. No. 3.8e-79;
 Matches 342; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 atgagctctatcggaacacttttcgaatcgcggaacctgcagacgatggcctcggg 60
 Db 1 atgagctctatcggaacacttttcgaatcgcggaacctgcagacgatggcctcggg 60
 QY 61 cctcagcaacacgaggaactccagcagctgccttcgctgctccgagcagcagctg 120
 Db 61 cctcagcaacacgaggaactccagcagctgccttcgctgctccgagcagcagctg 120
 QY 121 gatcagttgctccatgttctcatcatgatgctgcaacagacgagcagcagcgatgca 180
 Db 121 gatcagttgctccatgttctcatcatgatgctgcaacagacgagcagcagcgatgca 180
 QY 181 aatcaggagtggtggaacacacgacgacgagcgttcacaggaagccttgagtcggtg 240
 Db 181 aatcaggagtggtggaacacacgacgacgagcgttcacaggaagccttgagtcggtg 240
 QY 241 acgcagatgctgatgcagatcgtgatgcagctgatgcagacacgagcgcgcgcgcatg 300
 Db 241 acgcagatgctgatgcagatcgtgatgcagctgatgcagacacgagcgcgcgcgcatg 300
 QY 301 ggcggtggcggttcggttcacagcagcagcctggcgcggaacgcc 342
 Db 301 ggcggtggcggttcggttcacagcagcagcctggcgcggaacgcc 342

RESULT 2
 AAT85356

ID AAT85356 standard; DNA; 2004 BP.
 XX
 AC AAT85356;
 XX
 DT 09-DEC-1997 (first entry)
 XX
 DE Nephila clavipes spider silk protein 2 Kb DNA sequence.
 XX
 KW High strength film; fibre; woven article; parachutes; sails;
 KW absorber; body armour; heavy metal; biological weapon; chemical;
 KW flavour; fragrance; Nephila clavipes; ss.
 XX
 OS Nephila clavipes.
 XX
 FH Key Location/Qualifiers
 FT CDS 40...1980
 FT /*tag= a
 FT /product= silk_protein
 XX
 XX WO9708315-A1.
 PD 06-MAR-1997.
 XX
 XX 22-AUG-1996; 96WO-US13767.
 XX
 XX 22-AUG-1995; 95US-0517694.
 XX
 XX (BASE/) BASEL R M.
 XX (ELIO/) ELION G R.
 XX
 XX Basel RM, Elion GR;
 XX
 XX WPI: 1997-179272/16.
 DR P-PSDB; AAW271178.
 XX
 XX New opt. multimerised DNA sequences encoding spider silk protein -
 PT contg, both repetitive and non-repetitive sequences, useful for
 PT making high strength films, fibres, woven articles etc.
 XX
 XX Claim 14; Fig 1; 57pp; English.
 XX
 CC A process has been developed for the production of a DNA fragment
 CC encoding silk protein. The process involves: (a) selecting target DNA,
 CC from a silk-producing spider, that contains many repetitive and non-
 CC repetitive regions; (b) selecting a single-stranded DNA primer of at
 CC least 10 nucleotides with a sequence that is complementary to a region
 CC of the target; (c) repetitively combining the primer with melted target
 CC DNA, incubating the mixture with nucleotides and a DNA polymerase with
 CC proofreading activity to produce a DNA fragment which is complementary
 CC to the target and is at least 2 kb long. The present sequence
 CC represents a 2 Kb DNA sequence which encodes the spider silk protein
 CC from Nephila clavipes. The DNA fragment can be used to make fibres,
 CC films, woven articles, e.g. for use in parachutes, sails, body armour,
 CC and absorbers (e.g. of heavy metals, biological weapons, DNA, chemicals,
 CC flavours and fragrances). The high molecular weight (90-250 kb) of
 CC spider silk proteins can be produced on a commercial scale (at over
 CC 2 g/l cell mass). It has better tensile strength and elasticity than
 CC silkworm silk. Inclusion of both repetitive and non-repetitive regions
 CC ensures isolation of stable clones.
 XX
 XX Sequence 2004 BP; 481 A; 386 C; 791 G; 346 T; 0 other;

Query Match 11.3%; Score 38.8; DB 18; Length 2004;
 Best Local Similarity 50.0%; Pred. No. 1.6;
 Matches 97; Conservative 0; Mismatches 97; Indels 0; Gaps 0;
 QY 110 agcagcagctggatcagttgctgcctcatgttcacatgatgctgcaacagagccag 169
 Db 813 agcagcagctggaggtgctggaacagagagatcacggtcttctgggcaaggtgccc 872
 QY 170 gcagcagatgcaaatcagaggaggtgtggcaacgaacacgcgagacggtcaacaggaagcc 229

[illegible]

RESULT 9
AAT49316
ID AAT49316 standard; DNA; 1035 BP.

XX	Hypersensitive response; elicitor; <i>Pseudomonas solanacearum</i> ; plant;
KW	disease-resistance; <i>Escherichia coli</i> ; infiltration; virus;
KW	bacterium; fungus; pathogen; biological control agent; ss;

XX	
PD	19-DEC-1996.
XX	
PF	05-JUN-1996. 0650-17500010

XX
PI Beer SV, Wel Z;
XX

PT response elicitor polypeptide or protein
XX
PS Disclosure; Page 51-52; 69pp; English.
YY

infection, or plant infection with recombinant bacteria (non-infectious to the host plant, e.g. *Escherichia coli*) expressing the elicitor as a biological control agent, to allow

Sequence 1035 BP; 227 A; 330 C; 368 G; 110 T; 0 other;

211 aacggtcaacaggaaggtccctgagcttcgacgcagatgctgacgagatgcgtgatgcag 270
||| ||| ||||| ||| ||| ||||| ||||| |||
712 aggaagaccaggcgccctaccggcgctgctgcaaaagctgatgaagatcctgaacgcg 771

QY 331 ggcggcaacgcc 342
 ||
 Db 832 aaggggtgccggc 843

RESULT 10
 AAV54609
 ID AAV54609 standard; DNA; 1035 BP.

XX AC AAV54609;

XX DT 07-DEC-1998 (first entry)

XX DE Pseudomonas solanacearum hypersensitive response elicitor DNA.

XX KW Hypersensitive response elicitor; HRE; insect resistance;

XX KW biological control; transgenic plant; ss.

XX OS Pseudomonas solanacearum.

XX PN WO9837752-A1.

XX PD 03-SEP-1998.

XX PF 26-FEB-1998; 98WO-US03604.

XX PR 28-FEB-1997; 97US-0039226.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Wei Z, Zitter TA;

XX DR WPI: 1998-495374/42.

XX DR P-PSDB; AAW75865.

XX PT Use of hypersensitive response elicitor polypeptide - for

XX PT application to plants or seeds or transgenic plants or seeds for the

XX PT control of insects.

XX PS Disclosure; Page 15; 75pp; English.

XX CC This DNA sequence comprises an open reading frame that encodes a
 CC hypersensitive response elicitor (HRE) (see AAW75865 of Pseudomonas
 CC solanacearum. The invention relates to the use of a HRE
 CC polypeptide or protein to control insects on plants or plants grown
 CC from seed treated with HRE. Also claimed is a method of insect
 CC control for plants that involves: (a) providing a transgenic plant
 CC or seed transformed with a DNA molecule (see AAV54606-09) encoding a
 CC HRE polypeptide or protein (see AAW75862-67); and (b) growing the
 CC transgenic plants or transgenic plants produced from the transgenic
 CC seeds to control insects. HRE prevents direct insect damage to
 CC plants by feeding injury. It kills insects close to plants, and
 CC interferes with insect larval feeding on such plants. It also
 CC prevents insects from colonising host plants and releasing
 CC phytotoxins which result in disease damage to plants.

XX SQ Sequence 1035 BP; 227 A; 330 C; 368 G; 110 T; 0 other;

Query Match 11.0%; Score 37.6; DB 19; Length 1035;
 Best Local Similarity 55.3%; Pred. No. 2.7;
 Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 211 aacgggtcaacaggagcctgagtcctgttgacgcagatgctgacagatcgtgacg 270

Db 712 agcgaagaccaggcgccctaccggcgctgctgcaaaagctgatgaatcctgaacgcg 771

QY 271 ctgatgcgaacaccaggcgccgacgctgagcggtggttcgtcaacagcagcctg 330

Db 772 ctggtgcagatgatgcagcaagcgccctcgcgcgccgaacaggcgagggcggtcg 831

QY 331 ggcggcaacgcc 342
 ||
 ||

Db 832 aaggggtgccggc 843

RESULT 11

AAV39975

XX ID AAV39975 standard; DNA; 1035 BP.

XX AC AAV39975;

XX DT 09-NOV-1998 (first entry)

XX DE Pseudomonas solanacearum hypersensitive response elicitor DNA.

XX KW Hypersensitive response elicitor; HRE; growth; transgenic plant;

XX KW ss.

XX OS Pseudomonas solanacearum.

XX PN WO9832844-A1.

XX PD 30-JUL-1998.

XX PF 27-JAN-1998; 98WO-US01507.

XX PR 27-JAN-1997; 97US-0036048.

XX PA (CORR) CORNELL RES FOUND INC.

XX PI Beer SV, Qiu D, Wei Z;

XX DR WPI: 1998-427940/36.

XX DR P-PSDB; AAW62457.

XX PT Method for enhancing plant growth - comprises use of hypersensitive
 XX PT response elicitor polypeptide or protein which may also effect, e.g.
 XX PT increase in plant height or earlier germination seed

XX PS Disclosure; Page 21; 110pp; English.

XX CC This DNA sequence comprises an open reading frame that encodes a
 CC hypersensitive response elicitor (HRE) (see AAW62457) of Pseudomonas
 CC solanacearum. A method of enhancing growth in plants comprises:
 CC (a) applying a HRE polypeptide or protein in a non-infectious form
 CC to a plant or plant seed under conditions effective to enhance
 CC growth of the plant or plants grown from the seed, or (b) providing
 CC a transgenic plant or plant seed transformed with a DNA molecule
 CC encoding a HRE polypeptide or protein, and growing the transgenic
 CC plant or a plant produced from the transgenic seed under conditions
 CC effective to enhance plant growth. HRES (see AAW62454-59) or nucleic
 CC acids (see AAV62454-57) encoding them can be used to increase plant
 CC growth. The HRES may also result in increased plant height and
 CC yield, and effect early germination and maturation of plant seed
 CC and early colouration of fruit and plants.

XX SQ Sequence 1035 BP; 227 A; 330 C; 368 G; 110 T; 0 other;

Query Match 11.0%; Score 37.6; DB 19; Length 1035;
 Best Local Similarity 55.3%; Pred. No. 2.7;
 Matches 73; Conservative 0; Mismatches 59; Indels 0; Gaps 0;

QY 211 aacgggtcaacaggagcctgagtcctgttgacgcagatgctgacagatcgtgacg 270

Db 712 agcgaagaccaggcgccctaccggcgctgctgcaaaagctgatgaatcctgaacgcg 771

QY 271 ctgatgcgaacaccaggcgccgacgctgagcggtggttcgtcaacagcagcctg 330

Db 772 ctggtgcagatgatgcagcaagcgccctcgcgcgccgaacaggcgagggcggtcg 831

QY 331 ggcggcaacgcc 342
 ||
 ||

Db 832 aaggggtgccggc 843

RESULT	12	
AAV36430		
ID	AAV36430 standard; DNA; 1035 BP.	
XX		
AC	AAV36430;	
XX		
DT	26-OCT-1998 (first entry)	
XX		
DE	Hypersensitive response elicitor	
XX		
KW	Hypersensitive response elicitor	
KW	pathogen resistance; disease res.	
XX		
XX	<i>Pseudomonas syringae</i> .	
OS		
XX		
PN	W09824297-A1.	
XX		
PD	11-JUN-1998.	
XX		
PF	04-DEC-1997; 97WO-US22629.	
XX		
XX	05-DEC-1996; 96US-0033230.	
PR		
XX	(CORR) CORNELL RES. FOUND INC.	
PA		
XX		
PI	Beer SV, Qiu D, Wei Z;	
XX	WPI; 1998-332931/29.	
DR	P-PSDB; AAW61116.	
DR		
XX		
PT	Imparting pathogen resistance to	
PT	hypersensitive response elicitor	
XX		
PS	Disclosure; Page 24-25; 85pp; Eng	
XX		
CC	This nucleotide sequence from Pst	
CC	a coding region for a hypersensit	
CC	AAW61115). The invention relat	
CC	hypersensitive response induced	
CC	of seeds. Isolated HRE polypept	
CC	means of imparting pathogen resis	
CC	seeds. Alternatively, bacteria	
CC	HRE can be applied to the plant	
CC	containing a DNA molecule encodi	
CC	are used. HRE polypeptide sequen	
CC	Erwinia amylovora, Pseudomonas	
CC	Xanthomonas campestris pv. glyci	
CC	pelargonii (see AAW61113-18) are	
CC	pathogen resistance without using	
CC	environment or pathogenic to the	
CC	adjacent plants.	
XX		
XX		
SQ	Sequence 1035 BP; 227 A; 330 C; 3	

[illegible]

RESULT 13	
AAV83991	
ID	AAV83991 standard; DNA; 1035 BP.
XX	
AC	AAV83991;
XX	
DT	09-MAR-1999 (first entry)
XX	
DE	DNA encoding a hypersensitive response elicitor protein.
XX	
KW	Hypersensitive response elicitor protein; hairpin protein;
KW	disease resistance; seed quality; insect control; corn borer;
KW	Lepidoptera larvae; transgenic plant; ss.
XX	
OS	Pseudomonas solanacearum.
XX	
FX	Key
FX	Location/Qualifiers
FT	1..1035
FT	/*tag= a
FT	/product= Hypersensitive_response_elicitor_protein
XX	
PN	WO9854214-A2.
XX	
PD	03-DEC-1998.
XX	
PF	28-MAY-1998; 98WO-US10874.
XX	
PR	30-MAY-1997; 97US-0048109.
XX	
PA	(CORR) CORNELL RES FOUND INC.
PA	(EDEN-) EDEN BIOSCIENCE CORP..
XX	
PI	Beer SV, Laby RJ, Wei Z;
XX	
WPI;	1999-070210/06..
DR	P-PSDB; AAW87641.
XX	
PT	New fragments of an Erwinia hypersensitive response elicitor protein
PT	and related DNA - used to impart disease resistance to plants, to
PT	increase their growth and to control insects
XX	
PS	Disclosure; Page 16-17: 94pp; English.
XX	
CC	The present sequence encodes a hypersensitive response elicitor
CC	protein (also called hairpin protein) that is able to elicit a
CC	hypersensitive response in plants. The specification also describes
CC	hypersensitive response elicitors from other pathogenic organisms.
CC	The protein, in non-infectious form, is applied to plants to impart
CC	disease resistance (to a wide range of viral, bacterial and fungal
CC	pathogens), to improve growth (yield, quantity and quality of seed
CC	to provide earlier germination etc.) and to control insects (e.g. c
CC	borers, Lepidoptera larvae etc.) The same results are provided by
CC	transgenic plants expressing the protein.
XX	
SQ	Sequence 1035 BP; 227 A; 330 C; 368 G; 110 T; 0 other.

	Query Match	11.0%;	Score 37.6;	DB 20;	Length 1035;
	Best Local Similarity	55.3%;	Pred. No. 2.7;	Mismatches 0;	Gaps 0;
	Matches 73;	Conservative 0;	Mismatches 59;	Indels 0;	Gaps 0;
Qy	211	aacgggtcaaacagggaagcgctgagtccgttgtagcgcagatgctgatgcagatcgttatgcag	270		
Dd	712	agcgaagaccaggcgccctcacccggtgctgccaaagactgatgaagatcctgaacgcg	771		
Qy	271	ctgattgcagaaccacagggcgcgcggcatggcggttgcggttcggtccaacagcagacctg	330		
Dd	772	ctggtgcagatgatgcagcaaggcgcctcgccgcgcgaaccaaccaggcgcagggcggtctcg	831		
Qy	331	ggcggacaacgcc	342		
Dd	832	aagggtgccgcg	843		

Search completed: June 28, 2002, 11:25:04
Job time: 3736 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2002, 10:22:23 ; Search time 39.55 Seconds
(without alignments)
2124.062 Million cell updates/sec

Title: US-09-829-124-1
Perfect score: 342

Sequence: 1 atggactctatcggaacaa.....gcagcctggcggaagcc-342

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 303533 seqs, 122816752 residues

Total number of hits satisfying chosen parameters: 767066

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2_6/ptodata/2/ina/5A_COMB.seq.*
2: /cgn2_6/ptodata/2/ina/5B_COMB.seq.*
3: /cgn2_6/ptodata/2/ina/6A_COMB.seq.*
4: /cgn2_6/ptodata/2/ina/6B_COMB.seq.*
5: /cgn2_6/ptodata/2/ina/PTBUS_COMB.seq.*
6: /cgn2_6/ptodata/2/ina/Backfiles1.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	37.6	11.0	1035	1	US-08-891-254-8
2	37.6	11.0	1035	2	US-08-819-539-8
3	37.6	11.0	1035	2	US-09-030-270A-8
4	37.6	11.0	1035	4	US-08-984-207-8
5	37.6	11.0	1035	4	US-09-013-587-8
6	37.6	11.0	1035	5	PCT-US96-08819-8
7	37.2	10.9	2338	1	US-08-425-069-1
8	37.2	10.9	2338	2	US-08-317-844B-1
9	36.4	10.6	4403765	4	US-09-103-840A-2
10	36	10.5	657	4	US-08-949-155-5
11	36	10.5	1126	4	US-08-949-155-5
12	35.2	10.3	1665	3	US-08-881-784-8
13	35.2	10.3	1665	4	US-09-292-768-3
14	35.2	10.3	1665	4	US-09-292-768-3
15	35.2	10.3	1665	4	US-09-292-768-67
16	34.6	10.1	712	1	US-09-292-768-69
17	34.6	10.1	712	1	US-08-276-452A-72
18	34	9.9	543	6	US-08-798-744-72
19	33.8	9.9	1995	1	5273901-6
20	33.8	9.9	1995	1	US-08-425-069-3
21	33.8	9.9	1995	2	US-08-317-844B-3
22	33.4	9.8	557	4	US-09-056-556-221
23	33.2	9.7	397	3	US-09-253-691-3
24	33.2	9.7	3044	1	US-07-872-644-44
25	33.2	9.7	3044	1	US-08-297-494-44
26	33.2	9.7	3044	1	US-08-297-510-44
27	33.2	9.7	3044	1	US-08-479-532-44
28	33.2	9.7	3044	1	US-08-455-526-44

C	28	33.2	9.7	3044	1	US-08-455-525-44	Sequence 44, Appl
C	29	33.2	9.7	3044	3	US-09-139-491-44	Sequence 44, Appl
C	30	33.2	9.7	3044	5	PCT-US92-03222-44	Sequence 44, Appl
C	31	33.2	9.7	3468	1	US-07-951-715A-2	Sequence 2, Appl
C	32	33.2	9.7	3468	1	US-07-951-715A-4	Sequence 2, Appl
C	33	33.2	9.7	3468	2	US-08-459-448A-2	Sequence 2, Appl
C	34	33.2	9.7	3468	2	US-08-459-448A-4	Sequence 2, Appl
C	35	33.2	9.7	3468	3	US-08-459-595A-2	Sequence 4, Appl
C	36	33.2	9.7	3468	3	US-08-459-595A-4	Sequence 4, Appl
C	37	33.2	9.7	3468	3	US-08-459-504B-2	Sequence 2, Appl
C	38	33.2	9.7	3468	3	US-08-459-504B-4	Sequence 2, Appl
C	39	33.2	9.7	3468	3	US-08-459-444-2	Sequence 2, Appl
C	40	33.2	9.7	3468	3	US-08-459-444-4	Sequence 2, Appl
C	41	33.2	9.7	3468	3	US-09-053-549-3	Sequence 3, Appl
C	42	33.2	9.7	3468	3	US-09-053-549-5	Sequence 5, Appl
C	43	33.2	9.7	3468	4	US-09-547-422-2	Sequence 2, Appl
C	44	33.2	9.7	3468	4	US-09-547-422-4	Sequence 4, Appl
C	45	33	9.6	1515	4	US-09-292-768-5	Sequence 5, Appl

ALIGNMENTS

RESULT 1
US-08-891-254-8
Sequence 8, Application US/08891254
Patent No. 5776889
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Beer, Steven V.
TITLE OF INVENTION: Hypersensitive Response
TITLE OF INVENTION: Induced Resistance In Plants
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/891,254
FILING DATE: 10-JUL-1997
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/475,775
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael L.
REGISTRATION NUMBER: 30,727
REFERENCE/DOCKET NUMBER: 14603/10050
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 1035 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-891-254-8

Query Match 11.0%; Score 37.6; DB 1; Length 1035;
Best Local Similarity 55.3%; Pred. No. 0.19; Indels 0; Gaps 0;
Matches 73; Conservative 0; Mismatches 59;

RESULT 5
US-09-013-587-8
; Sequence 8, Application US/09013587
; Patent No. 6277814
; GENERAL INFORMATION:
; APPLICANT: Qiu, Dewen
; APPLICANT: Wei, Zhong-Min
; APPLICANT: Beer, Steven V.
; TITLE OF INVENTION: ENHANCEMENT OF GROWTH IN PLANTS
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: U.S.A.

```

1  RESULT 6
2  PCT-US96-08819-8
3  ; Sequence 8, Application PC/TUS9608819
4  ; GENERAL INFORMATION:
5  ; APPLICANT: Cornell Research Foundation, Inc.
6  ; TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
7  ; TITLE OF INVENTION: RESISTANCE IN PLANTS
8  ; NUMBER OF SEQUENCES: 9
9  ; CORRESPONDENCE ADDRESS:
10 ; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
11 ; STREET: Clinton Square, P.O. Box 1051
12 ; CITY: Rochester
13 ; STATE: New York
14 ; COUNTRY: U.S.A.
15 ; ZIP: 14603
16 ; COMPUTER READABLE FORM:
17 ; MEDIUM TYPE: Floppy disk
18 ; COMPUTER: IBM PC compatible
19 ; OPERATING SYSTEM: PC-DOS/MS-DOS
20 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
21 ; CURRENT APPLICATION DATA:
22 ; APPLICATION NUMBER: PCT/US96/08819
23 ; FILING DATE:
24 ; CLASSIFICATION:
25 ; PRIOR APPLICATION DATA:
26 ; APPLICATION NUMBER: US 08/475,775

```



```

Query Match      10.9%; Score 37.2; DB 2; Length 2338;
Best Local Similarity 49.5%; Pred. No. 0.33;
Matches 96; Conservative 0; Mismatches 98; Indels 0; Gaps 0;

QY 110 agcagcagctggatcagttgctgcgcacatgttcacatgatgatgctgcacacagagccagg 169
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 519 AGCAGCAGCGCGAGGAGTGCTGGACAAGCGGATACGGTGGTCTTGGTGGACAAGTGCGCG 578
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 170 gscagctgcaaatcatcaggagctggccaacgaacaaacgcagaaacggtcaacagagaagccc 229
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 579 ACAAGGAGGCTATGGAGGAGCTTGGAGCCAAAGGTGCTGGACGAGGAGGATTAGGTGGACA 638
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 230 tsggtccgttgacgcagatgctgatgcagatcgtgatgcagctgatgcagaacacagggcg 289
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 639 AGGTGCAGGTGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACGACG 698
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

QY 290 gcgcgcggcagtgggc 303
      | | | | |
Db 699 ACAAGGTGCTGGAC 712
      | | | | |

RESULT 9
US-09-103-840A-2/c
; Sequence 2, Application US/09103840A
; Patent No. 6294328
; GENERAL INFORMATION:
; APPLICANT: FLEISCHMAN, Robert D.
; APPLICANT: WHITE, Owen R.
; APPLICANT: FRASER, Claire M.
; APPLICANT: VENTER, John C.
; TITLE OF INVENTION: DNA SEQUENCES FOR STRAIN ANALYSIS IN MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS
; FILE REFERENCE: 24366-20007.00
; CURRENT APPLICATION NUMBER: US/09/103,840A
; CURRENT FILING DATE: 1998-06-24

```

```

: NAME: Weigs, J. Timothy
:
: REGISTRATION NUMBER: 38,241
: REFERENCE/DOCKET NUMBER: PF/5-30306/A/CGC1976
:
: TELECOMMUNICATION INFORMATION:
:
: TELEPHONE: 919-541-8587
:
: TELEFAX: 919-541-8689
:
: INFORMATION FOR SEQ ID NO: 1132:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 657 base pairs
:
: TYPE: nucleic acid
:
: STRANDEDNESS: single
:
: TOPOLOGY: linear
:
: MOLECULE TYPE: DNA (genomic)
:
: ORIGINAL SOURCE:
:
: ORGANISM: PAGI689UP
:
: US-08-998-416-1132

```

Query Match	10.5%	Score 36	DB 4	Length 657
Best Local Similarity	50.6%	Pred. No.	0.45	
Matches	87	Conservative	0	Mismatches 85; Indels 0; Gaps 0;
QY 109	gagcagcagctgcatcagttgctgcgcattgttcattcatgatgatgctgcaacagagccag	168		
Db 132	GAGNAGGAGCTGGCGATGCTGGACAAGCTGCTGGAGAAGACGACGGGGGGACGAAACCAG	191		
QY 169	ggcagcgatgcaaatcaggagtggtggcaacgaaacacccgacgaacggttcaacagagaagcc	228		
Db 192	CGCGTGAAGCAGCAGATGGTGTACTTGACGACNACTTCGCGCGTTCGGGGACCGAGGACG	251		
QY 229	ctgagtcgctgtgacgcagatgctgtagtcagatcgtgatgcagctgatgcaga	280		
Db 252	GTACGCGCGGTGACGACGTCGTGGCGCTCGACGGGACTGACCCGCTTCGAGA	303		

RESULT 11
US-89-949-155-5
: Sequence 5, Application US/08949155
: Patent No. 6271436
: GENERAL INFORMATION:
: APPLICANT: Piedrahita, Jorge A
: APPLICANT: Bazer, Fuller W
: TITLE OF INVENTION: Compositions and Methods for the
: TITLE OF INVENTION: Generation of Transgenic Animal Species
: NUMBER OF SEQUENCES: 51
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: ARNOLD, WHITE AND DURKEE
: STREET: P.O. Box 4433
: CITY: Houston
: STATE: TX
: COUNTRY: US
: ZIP: 77210-4433
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/949,155
: FILING DATE: Concurrently Herewith
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/027,338
: FILING DATE: 11-OCT-1996
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 60/046,094
: FILING DATE: 09-MAY-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Hibler, David W.
: REGISTRATION NUMBER: 41,071
: REFERENCE/DOCKET NUMBER: TANK:177
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (512) 418-3000

```

: TELEFAX: (713) 789-2679
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
:   LENGTH: 1126 base pairs
:   TYPE: nucleic acid
:   STRANDEDNESS: single
:   TOPOLOGY: linear
: FEATURE:
:   NAME/KEY: CDS
:   LOCATION: 51..1001
: US-08-949-155-5

```

Query Match 10.5%; Score 36; DB 4; Length 1126;
Best Local Similarity 49.0%; Pred. No. 0.54;
Matches 96; Conservative 0; Mismatches 100; Indels

QY	106	tccgagcagcgatgtgatctgtctgcacatttcatcatgatgatgctgcacacgagc	165
Db	231	TCTGACCAAGTGCAGGAGGAGCTGCTCAGCACCAAGCTACCCAGGAAGCTGACGGAGCTG	290
QY	166	caggggcgcgatgcaaatcaggagtgtggcaacgaacaaccgcagaaacggtcaacaggaa	225
Db	291	ATGAGGAGAGCATGAAGGAGGTGAAGGCCCTACCGGAGGAGTGTCGAGCGCGAGCTGGGC	350
QY	226	ggcctgagtccgttgacgcagatgctgatgcagatcgtgatgcagctgatgcagaaccag	285
Db	351	CCCGTAGCCACGAGGAGACGCAGCGGCCCTGTCCAAGGAGCTGCAGCGCGGCGAGGCCCG	410
QY	286	ggcggcgccggcatgg	301
Db	411	GTGGGCGCGACATGG	426

RESULT 12
US-08-881-784-8
; Sequence 8, Application US/08881784
; Patent No. 6083731
; GENERAL INFORMATION:
; APPLICANT: Croteau, Rodney B.
; APPLICANT: Lupien, Shari L.
; APPLICANT: Karp, Frank
; TITLE OF INVENTION: RECOMBINANT MATERIALS AND METHODS FOR
; THE PRODUCTION OF LIMONENE HYDROXYLASES
; NUMBER OF SEQUENCES: 58
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Christensen, O'Connor, Johnson and Kindness
; ADDRESSEE: PLLC
; STREET: 1420 Fifth Avenue, Suite 2800
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98101
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/881,784
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Shelton, Dennis K.
; REGISTRATION NUMBER: 26,997
; REFERENCE/DOCKET NUMBER: WSUR19777
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 224-0718
; TELEFAX: (206) 224-0779
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1665 base pairs
; TYPE: nucleic acid

Search completed: June 28, 2002, 11:27:18
Job time: 3895 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: June 28, 2002, 10:21:28 ; Search time 1603.91 seconds
(without alignments)
2877.941. Million cell updates/sec

Title: US-09-829-124-1
Perfect score: 342
Sequence: 1 atggactctatcggaacaa.....gcagcctggcggaacgcc 342

Scoring table: IDENTITY_NUC

Gap 10.0 ; Gapext 1.0

Searched: 13736207 seqs, 6748477542 residues

Total number of hits satisfying chosen parameters: 27472414

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

EST:*

1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_gss:*
13: em_gss_hum:*
14: em_gss_inv:*
15: em_gss_pln:*
16: em_gss_vit:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	43.8	12.8	526	12	AZ312601 1M0028104
2	41.8	12.2	537	10	BG930118
3	41.6	12.2	571	12	AQ919110
4	41.2	12.0	505	9	BB867074
5	41.2	12.0	968	12	CNS03F00
6	40.8	11.9	425	10	B1895704
7	40.6	11.9	803	12	AZ101123
8	40.6	11.9	325	10	BG929756
9	40.6	11.9	559	12	AZ661219
10	40.4	11.8	585	10	B1832289
11	40.4	11.8	815	9	AU142497
12	40.4	11.8	850	10	BG417526
13	40	11.7	291	10	BF592443
14	39.8	11.6	605	12	AZ640388
15	39.2	11.5	507	10	BF199083
16	39.2	11.5	650	12	AZ337339
17	39	11.4	420	10	BG562218

C 18	39	11.4	821	10	BF570046
19	38.8	11.3	546	9	AJ273961
20	38.8	11.3	650	10	BM488443
C 21	38.6	11.3	440	10	BF414279
C 22	38.6	11.3	718	12	AZ972907
23	38.4	11.2	1006	12	CNS04SSP
24	38.2	11.2	541	12	B73805
25	38	11.1	330	10	BG560855
26	38	11.1	354	9	BB870603
C 27	38	11.1	501	9	AT366107
28	38	11.1	615	9	AW106297
29	38	11.1	617	9	BB619432
30	38	11.1	635	9	BB619096
31	38	11.1	818	10	BG297454
32	37.8	11.1	362	10	BF072844
33	37.8	11.1	389	12	AZ537043
34	37.8	11.1	689	10	BG855860
35	37.8	11.1	812	12	CNS03005
36	37.8	11.1	1042	12	CNS04NIV
37	37.6	11.0	840	9	AL669315
38	37.4	10.9	609	10	B1633649
39	37.4	10.9	822	10	BG299472
40	37.2	10.9	261	12	BH326016
C 41	37.2	10.9	348	9	AI440942
42	37.2	10.9	423	10	BM305606
43	37.2	10.9	440	10	B1895386
C 44	37.2	10.9	472	9	AW756485
C 45	37.2	10.9	473	9	AW277347

ALIGNMENTS

RESULT 1
AZ312601/c
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
FEATURES

AZ312601 526 bp DNA linear GSS 29-SEP-2000
1M0028104R Mouse 10kb plasmid UUGC1M library Mus musculus genomic
clone UUGC1M0028104 R, DNA sequence.
AZ312601
GSS.
house mouse.
Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 526)
Dunn,D., Aoyagi,A., Barber,M., Beacorn,T., Duval,B., Hamil,C.,
Islam,H., Longacre,S., Mahmoud,M., Meenen,E., Pedersen,T., Reilly
,M., Rose,M., Rose,R., Stokes,R., Tingey,A., von Niederhausern,A.
and Wright,D., Weiss,R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0028 row: 1 column: 04
Seq primer: CACACAGGAACACGCTATGACC
Class: plasmid ends
High quality sequence stop: 526.
Location/Qualifiers
1..526
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="UUGC1M0028104"
/clone_lib="Mouse 10kb plasmid UUGC1M library"

FEATURES	SOURCE
1. The first two columns are labeled "FEATURES" and "SOURCE".	
2. The third column is labeled "FEATURES".	
3. The fourth column is labeled "SOURCE".	
4. The fifth column is labeled "FEATURES".	
5. The sixth column is labeled "SOURCE".	
6. The seventh column is labeled "FEATURES".	
7. The eighth column is labeled "SOURCE".	
8. The ninth column is labeled "FEATURES".	
9. The tenth column is labeled "SOURCE".	
10. The eleventh column is labeled "FEATURES".	
11. The twelfth column is labeled "SOURCE".	
12. The thirteenth column is labeled "FEATURES".	
13. The fourteenth column is labeled "SOURCE".	
14. The fifteenth column is labeled "FEATURES".	
15. The sixteenth column is labeled "SOURCE".	
16. The seventeenth column is labeled "FEATURES".	
17. The eighteenth column is labeled "SOURCE".	
18. The nineteenth column is labeled "FEATURES".	
19. The twentieth column is labeled "SOURCE".	
20. The twenty-first column is labeled "FEATURES".	
21. The twenty-second column is labeled "SOURCE".	
22. The twenty-third column is labeled "FEATURES".	
23. The twenty-fourth column is labeled "SOURCE".	
24. The twenty-fifth column is labeled "FEATURES".	
25. The twenty-sixth column is labeled "SOURCE".	
26. The twenty-seventh column is labeled "FEATURES".	
27. The twenty-eighth column is labeled "SOURCE".	
28. The twenty-ninth column is labeled "FEATURES".	
29. The thirtieth column is labeled "SOURCE".	
30. The thirty-first column is labeled "FEATURES".	
31. The thirty-second column is labeled "SOURCE".	
32. The thirty-third column is labeled "FEATURES".	
33. The thirty-fourth column is labeled "SOURCE".	
34. The thirty-fifth column is labeled "FEATURES".	
35. The thirty-sixth column is labeled "SOURCE".	
36. The thirty-seventh column is labeled "FEATURES".	
37. The thirty-eighth column is labeled "SOURCE".	
38. The thirty-ninth column is labeled "FEATURES".	
39. The fortieth column is labeled "SOURCE".	
40. The forty-first column is labeled "FEATURES".	
41. The forty-second column is labeled "SOURCE".	
42. The forty-third column is labeled "FEATURES".	
43. The forty-fourth column is labeled "SOURCE".	
44. The forty-fifth column is labeled "FEATURES".	
45. The forty-sixth column is labeled "SOURCE".	
46. The forty-seventh column is labeled "FEATURES".	
47. The forty-eighth column is labeled "SOURCE".	
48. The forty-ninth column is labeled "FEATURES".	
49. The fiftieth column is labeled "SOURCE".	
50. The fifty-first column is labeled "FEATURES".	
51. The fifty-second column is labeled "SOURCE".	
52. The fifty-third column is labeled "FEATURES".	
53. The fifty-fourth column is labeled "SOURCE".	
54. The fifty-fifth column is labeled "FEATURES".	
55. The fifty-sixth column is labeled "SOURCE".	
56. The fifty-seventh column is labeled "FEATURES".	
57. The fifty-eighth column is labeled "SOURCE".	
58. The fifty-ninth column is labeled "FEATURES".	
59. The sixtieth column is labeled "SOURCE".	
60. The sixty-first column is labeled "FEATURES".	
61. The sixty-second column is labeled "SOURCE".	
62. The sixty-third column is labeled "FEATURES".	
63. The sixty-fourth column is labeled "SOURCE".	
64. The sixty-fifth column is labeled "FEATURES".	
65. The sixty-sixth column is labeled "SOURCE".	
66. The sixty-seventh column is labeled "FEATURES".	
67. The sixty-eighth column is labeled "SOURCE".	
68. The sixty-ninth column is labeled "FEATURES".	
69. The seventieth column is labeled "SOURCE".	
70. The seventy-first column is labeled "FEATURES".	
71. The seventy-second column is labeled "SOURCE".	
72. The seventy-third column is labeled "FEATURES".	
73. The seventy-fourth column is labeled "SOURCE".	
74. The seventy-fifth column is labeled "FEATURES".	
75. The seventy-sixth column is labeled "SOURCE".	
76. The seventy-seventh column is labeled "FEATURES".	
77. The seventy-eighth column is labeled "SOURCE".	
78. The seventy-ninth column is labeled "FEATURES".	
79. The eightieth column is labeled "SOURCE".	
80. The eighty-first column is labeled "FEATURES".	
81. The eighty-second column is labeled "SOURCE".	
82. The eighty-third column is labeled "FEATURES".	
83. The eighty-fourth column is labeled "SOURCE".	
84. The eighty-fifth column is labeled "FEATURES".	
85. The eighty-sixth column is labeled "SOURCE".	
86. The eighty-seventh column is labeled "FEATURES".	
87. The eighty-eighth column is labeled "SOURCE".	
88. The eighty-ninth column is labeled "FEATURES".	
89. The ninetieth column is labeled "SOURCE".	
90. The ninety-first column is labeled "FEATURES".	
91. The ninety-second column is labeled "SOURCE".	
92. The ninety-third column is labeled "FEATURES".	
93. The ninety-fourth column is labeled "SOURCE".	
94. The ninety-fifth column is labeled "FEATURES".	
95. The ninety-sixth column is labeled "SOURCE".	
96. The ninety-seventh column is labeled "FEATURES".	
97. The ninety-eighth column is labeled "SOURCE".	
98. The ninety-ninth column is labeled "FEATURES".	
99. The hundredth column is labeled "SOURCE".	

```

1. SKI:
Location/Qualifiers
1. 537
/organism="Elmeria tenella"
/strain="Houghton"
/db_xref="taxon:5802"
/clone="etsHK077"
/clone_lib="Eth11"
/dev_stage="Sporozoite"
/lab_host="XLI-Blue MBF"
/note="Vector: Lambda ZAPII; Site_1: EcoRI; Site_2: XhoI; Sporozoites were excysted in vitro from E. tenella H oocysts, purified by column chromatography and mRNA extracted using a FASTTRACK kit (Invitrogen). cDNA was synthesised and a Uni-ZAP XR library was constructed using cDNA synthesis kit, ZAP-cDNA synthesis kit and ZAP-cDNA Gigapack III Gold cloning kit (Stratagene). The library was amplified once through E. coli XLI-Blue MBF."

```

[illegible]

Query Match	12.2%;	Score 41.8;	DB 10;	Length 537;
Best Local Similarity	48.2%;	Pred. No. 10;		
Matches 118;	Conservative 0;	Mismatches 127;	Indels 0;	Gaps 0;

[illegible]

RESULT	3
AQ919110	
LOCUS	571 bp DNA linear GSS 21-DEC-1999
DEFINITION	RPCI-23-275E22.TJ RPCI-23 Mus musculus genomic clone RPCI-23-275E22
ACCESSION	AQ919110
VERSION	AQ919110.1 GI:6608113

KEYWORDS	GSS
SOURCE	house mouse.
ORGANISM	Mus musculus
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 571)
AUTHORS	Zhao, S., Nierman, W., Feldblyum, T., Malek, J., Shatsman, S., Akinret B., Levinson, M., McGann, S., Tsegaye, G., Geer, K., Krol, M., de Jong, P. and Fraser, C.M.
TITLE	Mouse BAC End Sequences from Library RPCI-23
JOURNAL	Unpublished (1999)
COMMENT	Other_GSSs: RPCI-23-275E22.TV Contact: Shaying Zhao Department of Eukaryotic Genomics

TITLE	JOURNAL	COMMENT
1. The Role of the Teacher in the Classroom	Journal of Educational Research	1965, Vol. 68, No. 1, pp. 1-10
2. The Impact of Technology on Education	Journal of Educational Technology	1978, Vol. 1, No. 2, pp. 1-10
3. The Importance of Parental Involvement	Journal of Educational Research	1985, Vol. 88, No. 3, pp. 1-10
4. The Effect of Teacher Expectations on Student Achievement	Journal of Educational Research	1990, Vol. 93, No. 4, pp. 1-10
5. The Role of the School in the Community	Journal of Educational Research	1995, Vol. 98, No. 5, pp. 1-10
6. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2000, Vol. 103, No. 6, pp. 1-10
7. The Role of the Teacher in the 21st Century	Journal of Educational Research	2005, Vol. 108, No. 7, pp. 1-10
8. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2010, Vol. 113, No. 8, pp. 1-10
9. The Role of the Teacher in the 21st Century	Journal of Educational Research	2015, Vol. 118, No. 9, pp. 1-10
10. The Impact of Teacher Education on Student Achievement	Journal of Educational Research	2020, Vol. 123, No. 10, pp. 1-10

The Institute of Physical and
1-7-22 Suehiro-cho, Tsurumi-ku

-22 Suehiro-cho,

Email: genome-res@gsc.riken.go.jp.

genes. Genome Res. 10 (10), 1617-1630 (2000)
wagi, K., Fujiwake, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E.,
Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsuura
S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A. and
Havashizaki, Y.

sequencing pipeline with 384 m
10 (11), 1757-1771 (2000).

112 c 142 g 95 t

Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. *Genome Res.* 11 (2), 281-289 (2001). Please visit our web site (<http://genome.gsc.riken.go.jp>) for further details.

FEATURES	Location/Qualifiers
source	1. 505

1. 505

```

/organism="Mus musculus"
/db_xref="taxon:10090"
/clonename="G4D0005H22"
/clonelib="RIKEN full-length enriched, pooled cell lines"
/noted="pooled cell lines : (cell_line=CRL-1751 WEHI 164),
(cell_line=RCB-0035 WEHT-3),
(cell_line=RCB-2116 JC), (cell_line=RCB-0035 WEHT-3),
(cell_line=RCB-0464 Meth-A), (cell_line=RCB-0545 OHTA),
(cell_line=RCB-0559 K-1 Fl), (cell_line=RCB-1283 B16
melanoma), (cell_type=B cells, cell_line=CRL-1702 WEHI 231
), (cell_type=Leydig cells, cell_line=CRL-2065 MLTC-1),
(cell_type=Nullipotent stem cell, cell_line=CRL-2070 NE),
(tissue_type=bladder, cell_line=RCB-0544 MBT-2),
(tissue_type=bone marrow, cell_type=stroma cell,
cell_line= CRL-2028 SR-4987), (tissue_type=colon,
cell_line=RCB-0549 C1e-H3), (tissue_type=kidney,
cell_line=CCL-142 RAG), (tissue_type=submandibular gland,
cell_line=CRL-1734 SCA-9 clone 15), (strain=BALB/C,
cell_type=B cells, cell_line=CRL-1669 BCL1 clone 13.20-3B3
), (strain=C3H, tissue_type=brain, cell_line=CRL-1443

```

121 a	161 c	159 g	64 t
-------	-------	-------	------

Query Match	12.0%	Score 41.2;	DB 9;	Length 505;
Best Local Similarity	46.3%	Pred. NO. 14;		
Matches 136;	Conservative	0;	Mismatches 158;	Indels 0;
				Gaps 0;

38 acctgcagacgatggggcatcgggcctcagcaacacgaggaactccagccagctgcctt 97
51 AACTGCAGCAGCAGCTCTGGTCTCAAGCAGCAGCAGCAGCTCCGAAGACGCTCTGT 110

BASE COUNT	ORIGIN	cell_type=B cells, cell_line=CRL-1669 BCL1 Clone 13.20-3B3 (, (strain=C3H, tissue_type=brain, cell_line=CRL-1443 BC3H1)*
121 a	161 c	159 g 64 t

GI:17113284

38	acctgcagacatgggcacgcggccatgcagcaacagagactccagccagcagtcgcctt	97
51	AACTCGACGAGAGTCTCGTCTCAAGCAGCAGCAGCTCCAGAAGCAGCTCCTGT	110

Matches 136; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

Qy	117	gctggatcgttgctcgccatttcatcatgatgcgaacagagccaggcgca 176 : : : : : : : : : : :					
Db	289	GCCAGCAGGASCCMGACCSBHYYCCMGACCAGGAGCCAGMCCAASACCAGACCAGGA 230 : : : : : : : : : : :					
Qy	177	tgcaaatcagagtgtggcaacgaaacaaccgcagaacagtcgaacagga 224 					
Db	229	SCCAGACCGAGCCAGCACCAAGASCAGMCCACAASCCAGACCAGGA 182 					
RESULT	6						
B1895704							
LOCUS	B1895704	425 bp	mRNA	linear	EST 15-OCT-2001		
DEFINITION	EESTed23h03.y1 Eimeria tenella S5-2 cDNA Neg Selected Eimeria tenella cDNA 5', mRNA sequence.						
ACCESSION	B1895704						
VERSION	B1895704.1 GI:16138840						
KEYWORDS	EST.						
SOURCE	Eimeria tenella. Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae; Eimeria.						
REFERENCE	1. (bases 1 to 425)						
AUTHORS	Liberator,P., Diaz,C., Tang,X., Marra,M., Hillier,L., Kucaba,T., Martin,J., Wylie,T., Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person,B., Swallier,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter,E., Kohn,S., Florence.N., Shin,T., Jackson,Y., Cardenas,M., McCann,R., Waterston.R., Wilson,R. and Sibley,D. Washu-Merck Eimeria tenella project Unpublished (1999) Contact: David Sibley, Ph.D. Washu-Merck Eimeria tenella project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability. Seq primer: -4ORP from Gibco High quality sequence stop: 391. Location/Qualifiers 1..425 /organism="Eimeria tenella" /strain="LS18" /db_xref="taxon:5802" /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected" /dev_stage="Sporozoite stage" /lab_host="SOLR" /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>-10, from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminants."						
TITLE	Washu-Merck Eimeria tenella project						
JOURNAL	Unpublished (1999)						
COMMENT	Contact: David Sibley, Ph.D. Washu-Merck Eimeria tenella project Washington University School of Medicine 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA Tel: 314 286 1800 Fax: 314 286 1810 Email: est@watson.wustl.edu Contact David Sibley (toxoest@borcim.wustl.edu) for further information relating to organism, libraries, or clone availability. Seq primer: -4ORP from Gibco High quality sequence stop: 391. Location/Qualifiers 1..425 /organism="Eimeria tenella" /strain="LS18" /db_xref="taxon:5802" /clone_lib="Eimeria tenella S5-2 cDNA Neg Selected" /dev_stage="Sporozoite stage" /lab_host="SOLR" /note="Vector: Bluescript SK-; Site_1: EcoRI; Site_2: XhoI ; Sporozoites were obtained from in vitro sporulated and excysted oocysts of E.tenella grown in chickens. cDNA was synthesized from poly mRNA using an oligo-dT primer containing a XhoI site. Following second strand synthesis, EcoRI adapters were ligated to the cDNA and products were size-selected on sephacryl S500. The cDNA were ligated to EcoRI/XhoI prepared lambda ZapII (Stratagene). Clones were converted to phagemids by mass excision using ExAssist helper phage and E.coli SOLR cell (Stratagene). Clones were selected by negative hybridization against a pool of overrepresented ESTs (N>-10, from 1682 previous reads). Insert sizes range from 1.2-2.9Kb. The library may contain a small percentage of host or bacterial contaminants."						
FEATURES	source						
BASE COUNT	127 a	164 c	107 g	26 t	1 others		

```
Query Match      11.9%;   Score 40.8;   DB 10;   Length 425;
Best Local Similarity 51.3%;   Pred. No. 17;
Matches 119; Conservative 0; Mismatches 112; Indels 1; Gaps 1;
```

QY	10	atcggaacaacttttcgaatatatggccaacctgcagacgatgggcatcggggcctcagcaa	69
Db	12	AGCAGCAGCAGCAGTAAAGCAGCAGCAGCAGCAGCAGCAGTAAAGCAGAAAGCAGCAGAAG	71
QY	70	cacgaggaactccagccagcagtcgccttcggctggctccgagcagcagctggatcaattg	129
Db	72	CAGAAGCAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGGAGCAGCAGCAGCAGCAG	131
QY	130	ctcgccatgttcattcatgtatgctgcaacagcagccagggcagcagatgcataatcaggag	189
Db	132	CAGCAGCAGAAGCAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAAG	191
QY	190	tgtggcaacgaacaaccgcgagaacgctcaacagaaaggcctgagtccgttgacgcagatg	249
Db	192	CAGAACGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGAAG	251
QY	250	ctgatgcagatcgtgatgcagctgatgcagaaccag	285
Db	252	CAGAAGCAGAAGCAGAAGCAGAAGCAGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAG	287

RESULT	8
LOCUS	BG929756
DEFINITION	325 bp mRNA linear EST 31-DEC-2001 etsHST0177 <i>Eimeria tenella</i> cDNA clone etsHd066 5', mRNA sequence.
ACCESSION	BG929756
VERSION	BG929756.1
KEYWORDS	GI:18003146
SOURCE	EST.
ORGANISM	<i>Eimeria tenella</i> . <i>Eimeria tenella</i> . Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeriidae;

REFERENCE AUTHORS TITLE JOURNAL COMMENT	
Eukaryota; Alveolata; Apicomplexa; Coccidia; Eimeriida; Eimeria.	
1 (bases 1 to 325)	
Ng,S.T., Jangui,M.S., Shirley,M.W., Tomley,F.M. and Wan,K.L.	
Comparative EST analyses provide insights into gene expression in	
two asexual developmental stages of Eimeria tenella	
Unpublished (2001)	
Contact: Wan KL	
Centre for Gene Analysis and Technology	
Universiti Kebangsaan Malaysia	
43600 UKM Bangi, Selangor DE, Malaysia	
Tel: 6 03 8292997	
Fax: 6 03 8293249	
Email: klawan@pktrisc.cc.ukm.my	
PCR Primers	
FORWARD: T3	
BACKWARD: T7	

```

Seq primer: SK
FEATURES             Location/Qualifiers
     1..325
         /organism="Eimeria tenella"
         /strain="Houghton"
         /db_xref="taxon:5802"
         /clone="etsHD066"
         /clone_lib="EtH11"
         /dev_host="Sporozoite"
         /lab_host="XL1-Blue MRF".
     /note="vector: Lambda ZAPII; Site_1: ECORI; Site_2: XhoI;
Sporozoites were excysted in vitro from E. tenella H
oocysts, purified by column chromatography and mRNA
extracted using a FASTTRACK kit (Invitrogen). cDNA was
synthesised and a Uni-ZAP XR library was constructed.
using cDNA synthesis kit, ZAP-cDNA synthesis kit and
ZAP-cDNA Gigapack III Gold cloning kit (Stratagene). The
library was amplified once through E. coli XL1-Blue
MRF".

```

Query Match 11.9%; Score 40.6; DB 10; Length 325;
Best Local Similarity 51.5%; Pred. No. 17;

Matches 117: Conservative 0; Mismatches 109; Indels 1; Gaps 1;

Qy 110 agcagcagctgagtgatgctgcgcctggttcacatgatgatgctgcaaacagagccagg 169
|||||
Db 39 AGCAGCAGCGGAGCAGCAGCAGCTGCTGCTGCAGCAGCTACTGTCAGCAGTCCCTAG 98
|||||

Qy 170 gcagcgatgcaaatcagagtggtggaacgaacacccagcagacggttcaacaggaagcc 229
|||
Db 99 AGAGCAGCAGCAGCAGCAGCGGCGCACCACTGACCTGCAACAAAACAGCTGGTGC 158
|||||

Qy 230 tgagtccttgacgcagatgctgagtgatgctgagtgatgctgagtgaggaacccagg-gc 288
|||||
Db 159 AGCGTCAGTGGCGGAGCAGCAGCGGCGCACCACTGACCTGACCTGACCTGACCTGAC 218
|||||

Qy 289 ggcgcgcatggcggtggcggttcggtcgaacagacacccctggcg 335
|||||
Db 219 AGCACCAGCAGCAGCAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 265
|||||

RESULT 9
LOCUS AZ661219/c
DEFINITION IM0539M10R Mouse 10kb plasmid U06C1M library Mus musculus genomic
clone U06C1M0539M10 R, DNA sequence.
ACCESSION AZ661219
VERSION
KEYWORDS GSS.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 559)
Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly,
M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausern, A.
and Wright, D., Weiss, R.
Mouse whole genome scaffolding with paired end reads from 10kb
plasmid inserts
Unpublished (2000)
Contact: Robert B. Weiss
University of Utah Genome Center
University of Utah
Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
84112, USA
Tel: 801 585 5606
Fax: 801 585 7177
Email: ddunn@genetics.utah.edu
Insert Length: 10000 Std Error: 0.00
Plate: 0539 row: M column: 10
Seq primer: CACACAGGAACAGCTATGACC
Class: plasmid ends
High quality sequence stop: 559.
Location/Qualifiers
1. 559
/organism="Mus musculus"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="U06C1M0539M10"
/lab_host="Mouse 10kb plasmid U06C1M library"
/sex="Male"
/lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
/note="Vector: PWD42uv; Purified genomic DNA from M.
musculus C57BL/6J (male) was obtained from the Jackson
Laboratory Mouse DNA Resource
(http://www.jax.org/resources/documents/dnares/). The DNA
was hydrodynamically sheared by repeated passage through a
0.005 inch orifice at constant velocity. The sheared DNA
was blunt end-repaired with T4 DNA polymerase and T4
polynucleotide kinase. Adaptor oligonucleotides were
ligated to the blunt ends in high molar excess. The
adapted DNA was purified and size-selected for a 9.5 to
10.5 kb range using preparative agarose gel
electrophoresis. Vector DNA was prepared from a derivative

of pWD42 (gi|4732114|gb|AF129072.1), a copy-number
inducible derivative of plasmid R1. The vector was ligated
with adaptors complementary to the insert adaptors and
purified. The sheared, adapted mouse DNA was annealed to
adapted vector DNA, and transformed into
chemically-competent E. coli XL10-Gold (Stratagene) cells
and selected for ampicillin resistance."

BASE COUNT 17 a 250 c 143 g 149 t
ORIGIN

Query Match 11.9%; Score 40.6; DB 12; Length 559;
Best Local Similarity 51.1%; Pred. NO. 20;
Matches 120; Conservative 0; Mismatches 114; Indels 1; Gaps 1;

Qy 109 gacagcagctgagtgatgctgcgcctggttcacatgatgatgctgcaaacagagccagg 168
|||||
Db 423 GGGCAGCAGCAGCAGCGGCGCACCACTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 364
|||||

Qy 169 ggcagcgatgcaaatcagagtggtggaacgaacacccagcagacggttcaacaggaagcc 228
|||||
Db 363 GGGCAGCAGCAGCAGCGGCGCACCACTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 304
|||||

Qy 229 ctgagtccttgacgcagatgctgagtgatgctgagtgatgctgagtgaggaacccagg 287
|||||
Db 303 CAGGGCAGCAGCAGCGGCGCACCACTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 244
|||||

Qy 288 cgcgcgcatggcggtggcggttcggtcgaacagacacccctggcggaacgcc 342
|||||
Db 243 CAGCATCGGCAGCAGCAGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTGCGCTG 189
|||||

RESULT 10
BI832289
LOCUS 603080137F1 NIH_MGC_119 Homo sapiens cDNA clone IMAGE:5171817 5',
DEFINITION mRNA sequence.
ACCESSION BI832289
VERSION BI832289.1 GI:15943839
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 585)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Email: cgapbs@mail.nih.gov
Tissue Procurement: Life Technologies, Inc.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: L14M1427 row: j column: 10
High quality sequence start: 8
High quality sequence stop: 375.
Location/Qualifiers
1. 585
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:5171817"
/clone_lib="NIH_MGC_119"
/tissue_type="medulla"
/lab_host="DH10B"
/note="Organ: brain; Vector: pCMV-SPORT6; Site: 1; NotI;
Site 2: EcoRV (destroyed); RNA source normal medulla from
anonymous male age 27. Library is oligo-dT primed and
directionally cloned (EcoRV site is destroyed upon
cloning). Average insert size 1.3 kb, insert size range

[illegible]

Db 371 AGCACAAATGACCTTCTGAAGCTGAGAACCATGCTCATGAAAGTGGAGAATGAGGACATGA 430

Db 431 ATAAAGACCAGATCTTCTGGAAAAGAAGCTCCGCCGCATGCAAGAGATGATTG 490

491 CAAGCATGCAGGCCACAGATGCAGATGCAGATGCAGGGCGGGGATGGCGATGGGGGGGC 548

RESULT 12
BG417526

LOCUS BG417526 850 bp mRNA linear EST 23-OCT-2000
DEFINITION HVSMEK0017024f Hordeum vulgare testis/pericarp EST library

HVCdNA0013 (normal) Hordeum vulgare cDNA clone HVSMEk0017024f, m sequence

SEQUENCE:		
ACCESSION	BG417526	
VERSION	BG417526.1	CT_13233077

EXTENSION
 KEYWORDS
 EST.
 SOURCE
 00417520.1 01.13523077

SOURCE	ORGANISM
barley,	<i>Hordeum vulgare</i>

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooid

REFERENCE 1 (bases 1 to 850) ; Triticaceae; Hordeum.

AUTHORS
Wing, R., Close, T.J., Kleinbofs, A., Wise, R., Kannangara, G., von
Wetstein, D., Akhunov, E., Chin, A., Choi, D.W., Fenton R.D., Kianla

P., Otto, C., Simons, K., Zhang, D., Begum, D., Frisch, D., Yu, Y., Henry, D., Palmer, M., Rambo, T., Simmons, J., Oates, R. and Main, D.

TITLE
Development of a genetically and physically anchored EST resource for barley genomics: Morex testa/pericarp cDNA library

JOURNAL
Unpublished (2001)
COMMENT
Contact: Wing RA

Clemson University Genomics Institute
Clemson University

100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288

Fax: 864 656 4293
Email: rwings@clemonson.edu

Total hq bases = 528
Seq primer: AATTATCCCTCAGCTAAAGCC

High quality sequence stop: 833.
Location/Qualifiers

Source	Location/Qualifiers
1. .850	for action—negative influence

/organism= "horqueum vulgare"
/cultivar= "Morex"
/abundance= "4000000000"

```
/uo_xref="taxon:4513"
/clone="HVSMEK0017024f"
```

```
/clone_lib="Hordeum vulgare testa/pericarp EST library
HVCDNA0013 (normal)"
```

```
/lab_host="TJC121"  
/tissue_type="testa/pericarp"
```

/note=Vector: lambdaZAP; Site_1: EcoR1; Site_2: Xho1;
Plants were raised from seeds in a Controlled Environment

growth chamber maintained in continuous light at 18°C, and testa and pericarp were dissected from developing kernels.

at Washington State University, Pullman, WA (Kannangara, von Wetstein). Total RNA was prepared. poly(A) RNA was

purified, one cDNA library was made, and 1 million pfu were in vivo excised to give phagescript SK(-) cDNA

phagmids in the TJ Close lab at the University of California, Riverside (Athunov Chin Choi, close.athunov@ucr.edu).

Kilian, Otto, Simons, Zhang). Phagemids were plated and titered on *E. coli* (K12) strains (Khanov, Chan, Choi, Crose, Feilcke, Kozminski, Kuznetsov, and others) and phages were picked at the Gleason University Geomicrobiology Laboratory.

(Begum, Palmer, Frisch, Atkins and wing). Plasmid DNA preparations from recombinant and non-recombinant

performed at CUGI (Wing, Yu, Frisch, Henry, Simmons, Oat-

```

, Rambo, Main). The sequence has been trimmed to remove

```

vector sequence and contains a minimum of 100 bases of phred value 20 or above. For more details on library preparation and sequence analysis see <http://www.genome.clemson.edu/projects/barley>. To order this clone see <http://www.genome.clemson.edu/orders> Also see Close TJ, Wing R, Kleinhofs A, Wise R (2001) Genetically and physically anchored EST resources for barley genomics. Barley Genetics Newsletter 31:29-30. (<http://wheat.pw.usda.gov/ggpages/bgn/31/cover.html>)

```

BASE COUNT      249 a 104 c 230 g 187 t
ORIGIN

Query Match      11.8%; Score 40.4; DB 10; Length 850;
Best Local Similarity 51.1%; Pred. No. 25;
Matches 95; Conservative 0; Mismatches 91; Indels 0; Gaps 0;

QY 152 tgcgaacagagccgagcgatgcaaatcaggagtggtgcaacgaacccgcaga 211
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 177 TGAAGCCACAGAACGAGCCGCGTGCATTCCGAGTGGCGGAACGGCGCACACTCACA 236

QY 212 acggtcaacaggaagcctgagtccttgacgcagatgctgacgatcgtgatcgagc 271
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 237 GCCTGTGTCAGAGTGGACCCCGTTCACGTACAGGACCTGCAGGTCTATGACGAACA 296

QY 272 tgatgcagaacccagggcgccgcatggcggtggttcggtcgaacagcgcctgg 331
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 297 ACTTCAGACGGTCTCGCGCGGGAGGATTCCGCTCCGTGACGACGGCTTCTTGGCGG 356

QY 332 gcgcca 337
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 357 ATGGCA 362

```

```

RESULT 13
LOCUS    BF592443
DEFINITION 7156c04.x1 NCI-CGAP Br16 Homo sapiens cDNA clone IMAGE:3338694 3'
          similar to contains element MSRI MSRI repetitive element ;, mRNA
          sequence.
ACCESSION BF592443
VERSION    BF592443.1 GI:11684767
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 291)
            NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
            National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
            Unpublished (1997)
            Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-remail.nih.gov
            Tissue Procurement: Dennis Sgroi, M.D., Kristina Cole, M.D., Ph.D.
            student, Michael R. Emmert-Buck, M.D., Ph.D.
            cDNA Library Preparation: David B. Krizman, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone Distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL, send email to:
            info@image.llnl.gov
            Seq primer: -40RP from Gibco.
            Location/Qualifiers
                1..291
                /organism="Homo sapiens"
                /db_xref="taxon:9606"
                /clone="IMAGE:3338694"
                /clone_lib="NCI-CGAP_Br16"
                /sex="female"
                /tissue_type="lobular carcinoma in situ"
                /dev_stage="adult"
                /lab_host="DH10B"

```

/note="Organ: breast; Vector: pAMP1; mRNA made from breast carcinoma tissue, cDNA made by oligo-dT priming. Directionally cloned. Size-selected on agarose gel, average insert size 400 bp. Primary library, non-amplified."

```

BASE COUNT      86 a 43 c 139 g 23 t
ORIGIN

Query Match      11.7%; Score 40; DB 10; Length 291;
Best Local Similarity 51.7%; Pred. No. 23;
Matches 91; Conservative 0; Mismatches 85; Indels 0; Gaps 0;

QY 111 gcaacagctgagtcagttgctgccatgttcacatgatgatgctgaacagagccagg 170
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 69 GAAGCAGGAGGACGATCGGGAAGCAGGAGGACGATGGGGAGGAGGAGGACGAT 128

QY 171 -cagcgtgcaaatcaggagtggtgcaacgaacacccgagcaggtcaacaggaagcct 230
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 129 GGGGAGCAGGAGGACGATCGGGAAGCAGGAGGAGGACGATGGGGAGGAGGAGGACGA 188

QY 231 gactccgtgacgcagatgctgacgatcgtgatcgtgatcgtgatcgtgatcgtgatc 286
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 189 GATGCTGAACGAGGAGGACGATCGGAAGCAGGAGGAGGACGATGCGGAAGCAGG 244

```

```

RESULT 14
LOCUS    AZ640388
DEFINITION 1M0502E21F Mouse 10kb plasmid UUGC1M library Mus musculus genomic
          clone UUGC1M0502E21 F, DNA sequence.
ACCESSION AZ640388
VERSION    AZ640388.1 GI:11763788
KEYWORDS   GSS.
SOURCE     house mouse.
ORGANISM   Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
            1 (bases 1 to 605)
            Dunn, D., Aoyagi, A., Barber, M., Beacorn, T., Duval, B., Hamil, C.,
            Islam, H., Longacre, S., Mahmoud, M., Meenen, E., Pedersen, T., Reilly
            M., Rose, M., Rose, R., Stokes, R., Tingey, A., von Niederhausen, A.
            and Wright, D., Weiss, R.
            Mouse whole genome scaffolding with paired end reads from 10kb
            plasmid inserts
            Unpublished (2000)
            Contact: Robert B. Weiss
            University of Utah Genome Center
            University of Utah
            Rm. 308, Biomedical Polymers Research Bldg., 20 S. 2030 E., SLC, UT
            84112, USA
            Tel: 801 585 5606
            Fax: 801 585 7177
            Email: ddunn@genetics.utah.edu
            Insert Length: 10000 Std Error: 0.00
            Plate: 0502 row: E column: 21
            Seq primer: CGTTGTAACACGACGCCAGT
            High quality sequence stop: 605.
            Location/Qualifiers
                1..605
                /organism="Mus musculus"
                /strain="C57BL/6J"
                /db_xref="taxon:10090"
                /clone="UUGC1M0502E21"
                /clone_lib="Mouse 10kb plasmid UUGC1M library"
                /sex="Male"
                /lab_host="E. Coli strain XL10-Gold, T1-resistant, F-"
                /note="Vector: pMD42hv; Purified genomic DNA from M.
                musculus C57BL/6J (male) was obtained from the Jackson
                Laboratory Mouse DNA Resource
                (http://www.jax.org/resources/documents/dnares/). The DNA
                was hydrodynamically sheared by repeated passage through a

```

0.005 inch orifice at constant velocity. The sheared DNA was blunt end-repaired with T4 DNA polymerase and T4 polynucleotide kinase. Adaptor oligonucleotides were ligated to the blunt ends in high molar excess. The adapted DNA was purified and size-selected for a 9.5 to 10.5 kb range using preparative agarose gel electrophoresis. Vector DNA was prepared from a derivative of pWD42 (gi14732114|gb|AF129072.1); a copy-number inducible derivative of plasmid RI. The vector was ligated with adaptors complementary to the insert adaptors and purified. The sheared, adapted mouse DNA was annealed to adaptor vector DNA, and transformed into chemically-competent *E. coli* XL10-Gold (Stratagene) cells and selected for ampicillin resistance.

BASE COUNT 13 a 253 c 136 g 203 t
ORIGIN

Query Match 11.6%; Score 39.8; DB 12; Length 505;
Best Local Similarity 51.1%; Pred. No. 31;
Matches 118; Conservative 0; Mismatches 112; Indels 1; Gaps 1;
QY 110 agcagcagctggtatggtgtccatgttccatgatgatgtgcaatagagccagg 169
Db 483 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 424
QY 170 gcagcgtgcaaatcaggagtgtgcaacgaacacgcagacagcagcagcagc 229
Db 423 AGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 364
QY 230 tgagtcggtgacgcagatgctgagatcgtgagcagcagcagcagcagcagc 288
Db 363 AGGGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 304
QY 289 ggcgcgcgcagtggtggtggtggtggtggtggtggtggtggtggtggtggt 339
Db 303 AGAAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 253

RESULT 15
BF199083/c
LOCUS BF199083 507 bp mRNA linear EST 03-NOV-2000
DEFINITION 249143 MARC 2P1G Sus scrofa cDNA 5', mRNA sequence.
ACCESSION BF199083
VERSION BF199083.1 GI:11090834
KEYWORDS EST.
SOURCE pig.
ORGANISM Sus scrofa
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
1 (bases 1 to 507)
AUTHORS Fahrnenkrug,S.C., Freking,B.A., Rohrer,G.A., Smith,T.P.L., Casas,E.,
Stone,R.T., Heaton,M.P., Grosse,W.M., Bennett,G.A., Laegreid,W.W.
and Keele,J.W.
TITLE Design and use of two pooled tissue normalized cDNA libraries for
EST discovery in swine
JOURNAL Unpublished (2000)
COMMENT Contact: Smith TPL
USDA, ARS, US Meat Animal Research Center
PO Box 166, Clay Center, NE 68933-0166, USA
Tel: 402 762 4366
Fax: 402 762 4390
Email: smith@mail.marc.usda.gov
Single pass sequencing. Bases called and alt trimmed with phred
v0.980904.e. Vector identified by cross_match with the -minscore 18
and -minmatch 12 options.
PCR PRIMERS
FORWARD: AGGAACACGCTATGACCAT
BACKWARD: GTTTCCAGTCACGAGC
Plate: 81 row: N column: 2
Seq primer: ATTTAGGTGACACTATAG.
FEATURES
Location/Qualifiers
1..507
source

/organism="Sus scrofa"
/db_xref="taxon:9823"
/clone_lib="MARC 2P1G"
/tissue_type="pooled"
/lab_host="DH10B"
/note="Vector: pCMV SPORT6; Site_1: XbaI; Site_2: XhoI;
Library made from pooled tissue from testis, ovary,
endometrium, hypothalamus, pituitary, and placenta."
BASE COUNT 98 a 174 c 154 g 81 t
ORIGIN

Query Match 11.5%; Score 39.2; DB 10; Length 507;
Best Local Similarity 49.6%; Pred. No. 40;
Matches 127; Conservative 0; Mismatches 128; Indels 1; Gaps 1;
QY 52 ggcacggtggtcagcaacacagcagcagcagcagcagcagcagcagcagcagc 111
Db 367 GCCCACCCTCCGCGCAGTCTCTTTTACCACCTCACCACCTCAGTCCCTGGGGCGGCCTC 308
QY 112 cagcagcgtggtcagcagcagcagcagcagcagcagcagcagcagcagcagc 171
Db 307 CACCAGCTGGGTCAACCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 248
QY 172 agcagtgcaaatcaggagtgtgcaacgaacacgcagacagcagcagcagcagc 231
Db 247 CTGGGTCAAGCCTCTGCAGCTGGCGCTCATATCCCTCACAGCGGCACACAGCTCGACAC 188
QY 232 agtcggtgacgcagatgctgagatcgtgagcagcagcagcagcagcagcagcagc 291
Db 187 GG-CAGGCGGCTCAACAGCCATGACCCCTACTGTGGAGCTGGGGCCTCGCATGGGCCAA 129
QY 292 gccggcagtggtggtggtggtggtggtggtggtggtggtggtggtggtggtggt 307
Db 128 AGGGCGCTGGGCTCTG 113

Search completed: June 28, 2002, 10:50:23
Job time: 1735 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:16:03 ; Search time 23.71 Seconds
(without alignments)
534.054 Million cell updates/sec

Title: US-09-829-124-2

Perfect score: 584

Sequence: 1 MDSIGNFNSIGNLQTMGIG.....OGGAGMGGGGVNSSLGNA 114

Scoring table:

BLOSUM62

Gapop 10.0 ; Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

A_Geneseq_032802:*

- 1: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
- 2: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
- 3: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
- 4: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
- 5: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
- 6: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
- 7: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
- 8: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
- 9: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
- 10: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
- 11: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
- 12: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
- 13: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
- 14: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
- 15: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
- 16: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
- 17: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
- 18: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
- 19: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
- 20: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
- 21: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
- 22: /SIDS1/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	584	100.0	114	22 AAE10807	Xanthomonas campe
2	111	19.0	344	18 AAW6600	Hypersensitive res
3	111	19.0	344	19 AAW75865	Pseudomonas solana
4	111	19.0	344	19 AAW62457	Pseudomonas solana
5	111	19.0	344	19 AAW61116	Hypersensitive res
6	111	19.0	344	20 AAW87641	A hypersensitive r
7	111	19.0	344	21 AAW71099	Pseudomonas solana
8	111	19.0	344	21 AAY84860	A hypersensitive r
9	111	19.0	344	22 AAE06712	Pseudomonas solana
10	111	19.0	385	15 AAR45751	Erwinia amylovora
11	111	19.0	385	18 AAW06598	Hypersensitive res

12	111	19.0	403	19 AAW75863	Erwinia amylovora
13	111	19.0	403	19 AAW62455	Erwinia amylovora
14	111	19.0	403	19 AAW61114	Hypersensitive res
15	111	19.0	403	20 AAW87639	A hypersensitive r
16	111	19.0	403	21 AAW71093	Erwinia amylovora
17	111	19.0	403	21 AAY84854	A hypersensitive r
18	111	19.0	403	22 AAE06710	Erwinia amylovora
19	106	18.2	318	22 ABB68750	Drosophila melanog
20	96.5	16.5	1366	22 ABB68921	Drosophila melanog
21	96.5	16.5	1594	22 ABB63688	Drosophila melanog
22	94	16.1	20	19 AAW75867	Xanthomonas campe
23	94	16.1	20	19 AAW62459	Xanthomonas campe
24	94	16.1	20	19 AAW61118	Hypersensitive res
25	94	16.1	20	20 AAW87643	Hypersensitive r
26	94	16.1	20	21 AAY71101	Xanthomonas campe
27	94	16.1	20	21 AAY84862	Hypersensitive res
28	91.5	15.7	424	20 AAW75851	Hypersensitive res
29	91.5	15.7	424	21 AAY71098	Pseudomonas syring
30	91.5	15.7	424	21 AAY84859	A hypersensitive r
31	90	15.4	2328	22 ABB71657	Drosophila melanog
32	90	15.4	3190	22 ABB84634	Amino acid sequenc
33	90	15.4	3275	22 ABB70437	Drosophila melanog
34	88	15.1	800	16 AAR84882	Transcription fact
35	88	15.1	2414	16 AAR84882	Transcription fact
36	88	15.1	2414	19 AAW40057	Cellular transcrip
37	87	14.9	724	22 ABB65778	Drosophila melanog
38	87	14.9	1778	22 ABB60081	Drosophila melanog
39	85	14.6	147	22 ABB69847	Drosophila melanog
40	84.5	14.5	797	22 ABB62042	Drosophila melanog
41	84.5	14.5	915	22 ABB65275	Drosophila melanog
42	84	14.4	447	20 AAW96260	Hypersensitive res
43	84	14.4	447	21 AAY71094	Erwinia amylovora
44	84	14.4	447	21 AAY84855	A hypersensitive r
45	84	14.4	926	22 ABB65135	Drosophila melanog

ALIGNMENTS

RESULT 1

AAE10807
ID AAE10807 standard; Protein; 114 AA.

AC AAE10807;

DT 18-DEC-2001 (first entry)

DE Xanthomonas campestris hypersensitive response elicitor protein.

KW Plant pathogen; hypersensitive response elicitor; HRE; stress resistance;
KW disease resistance; Erwinia amylovora hypersensitive response elicitor;
KW harpin; signal transduction; growth enhancement; insect control;
KW virucide; fungicide; antibacterial.

OS Xanthomonas campestris.

XX WO200170988-A2.

PX 27-SEP-2001.

PF 19-MAR-2001; 2001WO-US08728.

XX 23-MAR-2000; 2000US-191649P.

PR 01-DEC-2000; 2000US-250710P.

XX (EDEN-) EDEN BIOSCIENCE CORP.

PI Song X, Fan H, Wei Z;

XX WPI; 2001-590177/66.

DR N-PSDB; AAD18127.

XX New plant pathogen hypersensitive response elicitor-receptor protein

PT isolated from plants, which upon silencing is used to study plant
PT signal transduction pathways leading to disease resistance and growth
XX enhancement
PS Disclosure; Page 18-19; 78pp; English.
XX
CC The present invention relates to an isolated protein which serves as a
CC receptor in plants for plant pathogen hypersensitive response elicitors
CC (HRE). The plants are made resistant to infection by viruses, bacteria
CC and fungi and are imparted with resistance against environmental stress
CC and insects through HRE treatment. The protein is useful for
CC understanding the harpin (Brwinia amylovora hypersensitive response
CC elicitor) induced signal transduction pathway in plants. The protein is
CC useful for studying the downstream components of signal transduction
CC pathway in plants which eventually leads to disease resistance, growth
CC enhancement, insect control and stress resistance. Imparting disease
CC resistance to plants through HRE treatment has the potential to treat
CC previously untreatable diseases, treating diseases systemically which
CC might not be treated separately due to cost and avoids the use of
CC infectious agents or environmentally harmful materials. By HRE treatment
CC enhanced plant growth is achieved which includes greater yield,
CC increased quantity of seeds produced, increased percentage of seeds
CC germinated, increased plant size, greater biomass, more and bigger
CC fruits, etc. which results in economic benefit to cultivators. Greater
CC yield, increased size and enhanced biomass allow greater revenue
CC generation from the given plot of plant. The present sequence is
CC Xanthomonas campestris pv. pelargonii hypersensitive response elicitor
CC protein.
XX
SQ Sequence 114 AA;

Query Match 100.08; Score 584; DB 22; Length 114;
Best Local Similarity 100.08; Pred. No. 4.8e-57;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDSIGNNFSNIGNLQTMGIGPOOHEDSSQSPSAGSEQLDQLAMFIMMQLQSQGSDA 60
|||||
Db 1 mdsignnfsnignlqtmgigpoohedsqqspssagseqldqlamfimmqlqsggsda 60
|||||

QY 61 NQPCNEPQNGQQGGLSPLTOMLMOIVMQLMONGGAGMGGSVNSLSLGGNA 114
|||||
Db 61 nqcneqngqngqgglspmtqlmqnlmqngggagmggsgvnsllgna 114
|||||

RESULT * 2
AAW06600
ID AAW06600 standard; Protein; 344 AA.
XX
AC AAW06600;
XX
XX 30-MAR-1997 (first entry)
XX
DE Hypersensitive response elicitor protein.
XX
XX Hypersensitive response; elicitor; Pseudomonas solanacearum; plant;
KW disease-resistance; Escherichia coli; infiltration; virus;
KW bacterium; fungus; pathogen; biological control agent.
XX
OS Pseudomonas solanacearum.
XX
XX WO9639802-A1.
XX
XX 19-DEC-1996.
XX
XX 05-JUN-1996; 96WO-US08819.
XX
XX 07-JUN-1995; 95US-0475775.
XX
XX (CORR) CORNELL RES FOUND INC.
XX
XX Beer SV, Wei 2;
XX

DR WPI: 1997-051614/05.
DR N-PSDB; AAT49316.
XX
PT Imparting pathogen resistance to plants - with hypersensitive
PT response elicitor polypeptide or protein
XX
PS Claim 13; Page 50-51; 69pp; English.
XX
CC This sequence represents a hypersensitive response elicitor from
CC Pseudomonas solanacearum. The elicitor may be used in a new method
CC for imparting pathogen resistance to plants, by application of the
CC elicitor in a non-infectious form to plant cells, by spraying,
CC injection, leaf abrasion, or plant infection with recombinant
CC bacteria (non-infectious to the host plant, e.g. Escherichia coli)
CC expressing the elicitor as a biological control agent, to allow
CC recombinant protein infiltration into the plant. The method
CC confers virus, bacterium or fungus disease-resistance on crops and
CC ornamental plants.
XX
SQ Sequence 344 AA;

Query Match 19.0%; Score 111; DB 18; Length 344;
Best Local Similarity 25.8%; Pred. No. 0.00045;
Matches 40; Conservative 17; Mismatches 54; Indels 44; Gaps 5;

QY 3 STGNFNSNIGNLQTM-----GIGPQOHEDSSQSPSAGSE-----QQLDQ 42
|||||
Db 58 sagngtngtnapakdgnanagandpskndpsksgapqskntgnvddannqdpmqalmq 117
|||||

QY 43 LLAMFIMMM-----LQSQGSDANQECG--NEQPQNGQQGGLSPLTOMLMOIVMQLMONG- 94
|||||
Db 118 lledivkilklaahmqpggndkngvggagkakagagggglaealqeieqilaqlgggg 177
|||||

QY 95 -----QGGAGMGGSVNSLSLGGN 113
|||||

Db 178 agagagagvggagagadggsgagagggangadggn 212
|||||

RESULT 3
AAW75865
ID AAW75865 standard; Protein; 344 AA.
XX
AC AAW75865;
XX
XX 07-DEC-1998 (first entry)
XX
DE Pseudomonas solanacearum hypersensitive response elicitor (HRE).
XX
KW Hypersensitive response elicitor; HRE; insect resistance;
KW biological control; transgenic plant.
XX
OS Pseudomonas solanacearum.
XX
XX WO9837752-A1.
XX
XX 03-SEP-1998.
XX
XX 26-FEB-1998; 98WO-US03604.
XX
XX 28-FEB-1997; 97US-0039226.
XX
XX (CORR) CORNELL RES FOUND INC.
XX
XX Wei 2, Zitter TA;
XX
XX WPI: 1998-495374/42.
XX
XX N-PSDB; AAV54609.
XX
XX Use of hypersensitive response elicitor polypeptide - for
XX application to plants or seeds or transgenic plants or seeds for the
XX control of insects.

This is the deduced amino acid sequence of the hypersensitive

Best Local Similarity 25.8%; Pred. No. 0.00045;
Matches 40; Conservative 17; Mismatches 54; Indels 44; Gaps 5;

QY 3 SIGNFSNIGNLQTM-----GIGPQOHHSSQQSPSAGSE-----QQLDQ 42
DB 58 sagngtntgnapakdgnanagandpskndpskqpsanktgnvddannqpmqalmq 117
QY 43 LLAMFIMM-----LQSSGSDANQECG--NEQPQNGQOEGLSPLTQMLQIVMQLMQN- 94
DB 118 lledlvkllkaalhmqqpgndkngvggagangakagagggglaealqeilaqlg99g 177
QY 95 -----QGGAGMGGGSVNSLGGN 113
DB 178 agaggaggvggagadgsgaggaggaggadggn 212

RESULT 8

AAE06712
ID AAE06712 standard; Protein; 344 AA.

AC AAE06712;

XX 08-AUG-2000 (first entry)

DT A hypersensitive response elicitor protein.

DE Hypersensitive response; insect control; disease resistance;
KW Hypersensitive response elicitor; plant growth; vegetable; crop;
KW ornamental plant.

XX Pseudomonas solanacearum.

XX WO200020452-A2.

XX 13-APR-2000.

XX 05-OCT-1999; 99WO-US23181.

XX 05-OCT-1998; 98US-0103050.

XX (EDEN-) EDEN BIOSCIENCE CORP.

XX Wei 2, Fan H, Niggemeyer JL;

XX WPI; 2000-303745/26.

XX N-PSDB; AAA14944.

XX Hypersensitive response elicitor polypeptides useful for imparting
PT enhanced growth, disease resistance and insect resistance to plants,
PT especially vegetables and ornamental flowers

XX Disclosure; Page 28-29; 100pp; English.

XX The present sequence represents a hypersensitive response elicitor
XX polypeptide. The specification describes hypersensitive response
XX elicitor polypeptide fragments, which do not elicit a hypersensitive
XX response. Instead, the proteins impart disease resistance to plants,
XX enhance plant growth, and/or control insects. The polypeptide
XX fragments may be used to these properties to plants. The plants which
XX may be treated in this way include vegetables, crops and ornamental
XX plants such as alfalfa, rice, wheat, barley, rye, cotton, sunflower,
XX peanut, corn, potato, sweet potato, bean, pea, chickory, lettuce,
XX endive, cabbage, brussel sprout, beet, parsnip, turnip, cauliflower,
XX broccoli, radish, spinach, onion, garlic, eggplant, pepper, celery,
XX carrot, squash, pumpkin, zucchini, cucumber, apple, pear, melon,
XX citrus, strawberry, grape, raspberry, pineapple, soybean, tobacco,
XX tomato, sorghum or sugarcane, Arabidopsis thaliana, Saintpaulia,
XX petunia, pelargonium, poinsettia, chrysanthemum, carnation or zinnia.

XX Sequence 344 AA;

Query Match

19.0%; Score 111; DB 21; Length 344;

Best Local Similarity 25.8%; Pred. No. 0.00045;
Matches 40; Conservative 17; Mismatches 54; Indels 44; Gaps 5;

QY 3 SIGNFSNIGNLQTM-----GIGPQOHHSSQQSPSAGSE-----QQLDQ 42
DB 58 sagngtntgnapakdgnanagandpskndpskqpsanktgnvddannqpmqalmq 117
QY 43 LLAMFIMM-----LQSSGSDANQECG--NEQPQNGQOEGLSPLTQMLQIVMQLMQN- 94
DB 118 lledlvkllkaalhmqqpgndkngvggagangakagagggglaealqeilaqlg99g 177
QY 95 -----QGGAGMGGGSVNSLGGN 113
DB 178 agaggaggvggagadgsgaggaggaggadggn 212

RESULT 9

AAE06712
ID AAE06712 standard; Protein; 344 AA.

AC AAE06712;

XX 16-OCT-2001 (first entry)

DT Pseudomonas solanacearum hypersensitive response elicitor protein.

DE Hypersensitive response elicitor; oomycete; transgenic plant; infection;
KW gene therapy; crop loss; antifungal.

XX Pseudomonas solanacearum.

XX WO200155347-A1.

XX 02-AUG-2001.

XX 26-JAN-2001; 2001WO-US02579.

XX 26-JAN-2000; 2000US-0178565.

XX (CORR) CORNELL RES FOUND INC.

XX Beer SV, Bauer DW;

XX WPI; 2001-488791/53.

XX N-PSDB; AAD12808.

XX New chimeric gene, useful for controlling plant-pathogenic fungi and
PT producing oomycete-resistant transgenic plants, comprises first DNA
PT encoding hypersensitive response elicitor, promoter and regulatory
PT region

XX Claim 20; Page 16-17; 72pp; English.

XX The invention relates to a chimeric gene that includes a first DNA
XX molecule encoding a hypersensitive response elicitor protein or
XX polypeptide, promoter operably linked 5' to the first DNA molecule
XX to induce transcription of the first DNA molecule in response to
XX activation of the promoter by an oomycete and a 3' regulatory region
XX operably linked to the first DNA molecule. The invention also relates
XX to a transgenic plant resistant to disease resulting from oomycete
XX infection, the transgenic plant including the chimeric gene, wherein
XX the promoter induces transcription of the first DNA molecule in
XX response to infection of the plant by an oomycete. The chimeric gene
XX is used in gene therapy. The chimeric gene is useful as an effective
XX and safe means of controlling plant-pathogenic fungi, particularly
XX oomycetes, which are responsible for major crop loss and is also useful
XX for producing transgenic plants of the invention. The present sequence
XX is Pseudomonas solanacearum hypersensitive response elicitor protein.

XX Sequence 344 AA;

Query Match

19.0%; Score 111; DB 22; Length 344;

Best Local Similarity 25.8%; Pred. No. 0.00045;
Matches 40; Conservative 17; Mismatches 54; Indels 44; Gaps 5;

QY 3 SIGNFNIGNLQTM-----GIGPOQHEHSSQSQSPSAGSE-----QQLDQ 42
Db 58 sagntntgnapakdnaganagandpskndpsksqapqsanktgnvddannqdpmqalmq 117
QY 43 LLAMFIMM-----LQSQSGSDANQEGC--NEQPQNGQOGLSPLTQMLMIVMQLMQN- 94
Db 118 lledlvklkaalhmqqpgndkngvggagkakgagggglaealqeieqlaqlggg 177
QY 95 -----QGGAGMGGGGSVNSSLGN 113
Db 178 agaggaggvggagdgadgsgagaggagadggn 212

RESULT 10
AAW45751
ID AAR45751 standard; Protein; 385 AA.
XX AC AAR45751;
XX DT 11-JUL-1994 (first entry)
XX DE Erwinia amylovora harpin.
XX KW Harpin; hypersensitive response elicitor; HR-elicitor; fire blight;
XX KW Rosaceae; apple; pear; phytopathogenic bacteria; defence reaction;
XX KW hypersensitive reaction and pathogenicity; hrpN; gene cluster; ds.
XX OS Erwinia amylovora.
XX FH Key Location/Qualifiers
FT Misc-difference 372 /note= "corresponds to CAG codon"
FT Misc-difference 373 /note= "corresponds to CAC codon"
XX PN WO9401546-A.
XX PD 20-JAN-1994.
XX PF 30-JUN-1993; 93WO-0506243.
XX PR 01-JUL-1992; 92US-0907935.
XX PA (CORR) CORNELL RES FOUND INC.
XX PI Bauer DW, Beer SV, Collmer A, He S, Laby R, Wei Z;
XX DR WPI; 1994-035054/04.
XX DR N-PSDB; AAQ55751.
XX PT Hypersensitive response elicitor protein derived from Erwinia
XX PT amylovora - and DNA encoding it, useful for developing harpin
XX PT inhibitors to prevent e.g. fire blight of fruit
XX PS Claim 6; Page 27-28; 47pp; English.
XX CC The hrpN gene was isolated from E.amylovora using a 48-fold
XX CC degenerate oligonucleotide probe corresponding to amino acids 9-15
XX CC at the N-terminus of harpin. The 44kD protein encoded by the hrpN
XX CC gene is a hypersensitive response elicitor protein. The harpin is
XX CC thought to be an archetype for HR elicitors from phytopathogenic
XX CC bacteria.

Query Match 19.0%; Score 111; DB 15; Length 385;
Best Local Similarity 31.5%; Pred. No. 0.00052;
Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;

QY 5 GNNFNSIGNL---OTMGI-GPOQHEHSSQSQSPS-AGSEQQLDQLLAMFIMMLOQS---Q 56
Db 118 gnttstnspldqalgnstsqnddstsgtdstsdspmqgllkmfseim--qslfgd 175
QY 57 GSDANQ--ECGNEQPQNGQOQ-----EGLSPLTQMLMIVMQLMQNQGAGMGGGGSVNSL 110
Db 176 gqdggtqgsssggkqpteqeqnaykkgvtdalsg lmgnglsqllngpglggggggnagtg 235
QY 111 CGNA 114
Db 236 dgss 239

RESULT 11
AAW06598
ID AAW06598 standard; Protein; 385 AA.
XX AC AAW06598;
XX DT 30-MAR-1997 (first entry)
XX DE Hypersensitive response elicitor protein.
XX KW Hypersensitive response; elicitor; Erwinia amylovora; plant;
XX KW disease-resistance; Escherichia coli; infiltration; virus;
XX KW bacterium; fungus; pathogen; biological control agent.
XX OS Erwinia amylovora.
XX PN WO9639802-A1.
XX PD 19-DEC-1996.
XX PF 05-JUN-1996; 96WO-US08819.
XX PR 07-JUN-1995; 95US-0475775.
XX PA (CORR) CORNELL RES FOUND INC.
XX PI Beer SV, Wei Z;
XX DR WPI; 1997-051614/05.
XX DR N-PSDB; AAT49314.
XX PT Imparting pathogen resistance to plants - with hypersensitive
XX PT response elicitor polypeptide or protein
XX PS Claim 7; Page 46-47; 69pp; English.
XX CC This sequence represents a hypersensitive response elicitor from
XX CC Erwinia amylovora, with a mol.wt. of 37,000. The elicitor has a pI
XX CC of 4.3, thermostability at 100 deg C for at least 10 min, and
XX CC contains no cysteine. The elicitor may be used in a new method for
XX CC imparting pathogen resistance to plants, by application of the
XX CC elicitor in a non-infectious form to plant cells, by spraying,
XX CC injection, leaf abrasion, or plant infection with recombinant
XX CC bacteria (non-infectious to the host plant, e.g. Escherichia coli)
XX CC expressing the elicitor as a biological control agent, to allow
XX CC recombinant protein infiltration into the plant. The method
XX CC confers virus, bacterium or fungus disease-resistance on crops and
XX CC ornamental plants.

Query Match 19.0%; Score 111; DB 18; Length 385;
Best Local Similarity 31.5%; Pred. No. 0.00052;
Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;

QY 5 GNNFNSIGNL---OTMGI-GPOQHEHSSQSQSPS-AGSEQQLDQLLAMFIMMLOQS---Q 56
Db 118 gnttstnspldqalgnstsqnddstsgtdstsdspmqgllkmfseim--qslfgd 175

Db 176 gddgtqsssggkqptegcnaaykkvtdalslmgnglsqllgngglgagggnagtgl 235

QY 111 GGNA 114
|::

Db 236 dqss 239

AAW62455	
ID	AAW62455 standard; Protein; 403 AA.
XX	
AC	AAW62455;
XX	
DT	09-NOV-1998 (first entry)
XX	
DE	Erwinia amylovora hypersensitive response elicitor (HRE).
XX	
KW	Hypersensitive response elicitor; HRE; growth; transgenic plant.
XX	
OS	Erwinia amylovora.
XX	
PN	WO9832844-A1.
XX	
PD	30-JUL-1998.
XX	
PF	27-JAN-1998; 98WO-US01507.
XX	
PR	27-JAN-1997; 97US-0036048.
XX	
PA	(CORR) CORNELL RES FOUND INC.
XX	
PI	Beer SV, Qiu D, Wei Z;
XX	
WPI	1998-427940/36.
XX	
N-PSDB	AAV39973.
XX	
PT	Method for enhancing plant growth - comprises use of hypersensitive
PT	response elicitor polypeptide or protein which may also effect, e.g.
PT	increase in plant height or earlier germination seed
PS	Disclosure; Page 15-16; 110pp; English.

This is the deduced amino acid sequence of the 39 kDa, heat stable hypersensitive response elicitor (HRE) of *Erwinia amylovora*. A method of enhancing growth in plants comprises: (a) applying a HRE polypeptide or protein in a non-infectious form to a plant or plant seed under conditions effective to enhance growth of the plant or plants grown from the seed, or (b) providing a transgenic plant or plant seed transformed with a DNA molecule encoding a HRE polypeptide or protein, and growing the transgenic plant or a plant produced from the transgenic seed under conditions effective to enhance plant growth. HRES (see AAMG2454-59) or nucleic acids encoding them (see AAV39973-75) can be used to increase plant growth. The HRES may also result in increased plant height and yield, and effect early germination and maturation of plant seed and early colouration of fruit and plants. E. amylovora HRE can be applied to tomato plants to enhance growth without causing disease in that species; this bacterium is a pathogen of apple and pear but not of tomato.

Sequence 403 AA:

Query Match 19.0%; Score 111; DB 19; Length 403;
Best Local Similarity 31.5%; Pred. No. 0.00055;
Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;

```
QY      5  GNNFNSIGNL---QTGMI-GPOHEDSSQSPS-AGSEQLDOLLAMFTMMMLAQSS--Q   56
      ||| :: |::||| :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||
Db     118 gnnatttspdqalngstqndtsdgtstdsdssdpmqgllkmfseim--qsifgd 175
      |||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||
QY     57 GSDANO--ECGNPQPONGQQ----EGLSLPTQLMLQJIVMQLNQGNGMGGSVNSSL 110
      |||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||:: :|||
```

Db	176	gqdcqtgsssggkqptegeqnaykqvtdalsqlmgnglsqllgngglgqgqgnaegtgl	233
Qy	111	GGNA 114	
Db	236	dgss 239	
RESULT	14		
AAW61114			
ID	AAW61114	standard; Protein; 403 AA.	
XX	AC		
XX	AC		
XX	XX		
DT	26-OCT-1998	(first entry)	
XX	XX		
DE	XX	Hypersensitive response elicitor protein (39 kDa).	
DE	XX		
KW	XX	Hypersensitive response elicitor; transgenic plant; seed;	
KW	XX	pathogen resistance; disease resistance; crop protection.	
XX	OS		
XX	XX	Erwinia amylovora.	
XX	XX		
PN	WO9824297-A1.		
XX	XX		
PD	11-JUN-1998.		
XX	XX		
PF	04-DEC-1997;	97WO-US22629.	
XX	XX		
PR	05-DEC-1996;	96US-0033230.	
XX	XX		
PA	(CORR)	CORNELL RES FOUND INC.	
XX	XX		
PI	Beer SV, Qiu D, Wei Z;		
XX	XX		
WPI:	1998-332931/29.		
DR	DR	N-PSDB; AAV36428.	
XX	XX		
PT	Imparting pathogen resistance to plants - by applying a		
PT	hypersensitive response elicitor polypeptide to seeds		
XX	XX		
PS	Disclosure; Page 18-20; 85pp; English.		
XX	XX		
CC	XX	This is the 39 kDa hypersensitive elicitor (HRE) protein of	
CC	CC	Erwinia amylovora. It is heat stable at 100 degC for at least 10	
CC	CC	min, has a pI of approximately 4.3, and contains substantially no	
CC	CC	cysteine. The invention relates to methods of imparting	
CC	CC	hypersensitive response induced resistance to plants by treatment	
CC	CC	of seeds. Isolated HRE proteins can be applied to seeds as a means	
CC	CC	of imparting pathogen resistance to plants grown from the seeds.	
CC	CC	Alternatively, bacteria containing the gene encoding the HRE can be	
CC	CC	applied to the plant seeds, or transgenic plant seeds containing a	
CC	CC	DNA molecule encoding an HRE polypeptide or protein are used. HRE	
CC	CC	polypeptide sequences from Erwinia chrysanthemi, Erwinia amylovora,	
CC	CC	Pseudomonas syringae, Pseudomonas solanacearum, Xanthomonas	
CC	CC	campestris pv. glycines and Xanthomonas campestris pathogenii (see	
CC	CC	AAW61113-18) are provided. The methods can impart pathogen	
CC	CC	resistance without using agents which are harmful to the	
CC	CC	environment or pathogenic to the plant seed being treated, or to	
CC	CC	adjacent plants. E. amylovora causes disease in apple or pear	
CC	CC	but not tomato. However, it elicits a hypersensitive response in	
CC	CC	tomato. Thus, E. amylovora can be applied to tomato seeds to	
CC	CC	impart pathogen resistance without causing diseases in plants of	
CC	CC	that species.	
XX	XX		
Sequence	403 AA;		
XX	XX		
SO	SO		

```

Query Match      19.0%  Score 111;  DB 19;  Length 403;
Best Local Similarity 31.5%  Pred. No. 0.00055;
Matches 39;  Conservative 21;  Mismatches 48;  Indels 16;  Gaps 7;

QY      5  GNNPNTGNTGNL---QTMGI-GPQQIHEDSSQQSPS-AGSEQQLDQLLMTFMMLQOS---Q.56
      |||  ::  |  -  |||  :  ::  ::  :  :  :  :  |||  ||  :  ||

```

Db 118 gntttstnspldgalinstsqnddstsgtdstsdssdpmqlllkmfseim--qslfgd 175

QY 57 GSDAMQ--ECGNEOPONGQO---EGLSPLTQMLQIVQMOLMONOGGACMGGGGVNSSL 110

Db 176 qgdgtqgsssgkpteqeqaykkgvtdalsgimgngisqllngdg199ggggnagtgl 235

QY 111 GGN 114

Db 236 dgss 239

RESULT .15

AAW87639

ID AAW87639 standard; Protein; 403 AA.

XX

AC AAW87639;

XX

DT 09-MAR-1999 (first entry)

XX

DE A hypersensitive response elicitor protein.

XX

KW Hypersensitive response elicitor protein; hairpin protein;

KW disease resistance; seed quality; insect control; corn borer;

KW Lepidoptera larvae; transgenic plant.

XX

OS Erwinia amylovora.

XX

PN WO9854214-A2.

XX

PD 03-DEC-1998.

XX

PF 28-MAY-1998; 98WO-US10874.

XX

PR 30-MAY-1997; 97US-0048109.

XX

PA (CORR) CORNELL RES FOUND INC.

PA (EDEN-) EDEN BIOSCIENCE CORP.

XX

PI Beer SV, Laby RJ, Wei Z;

XX

DR WPI; 1999-070210/06.

DR N-PSDB; AAV83989.

XX

PT New fragments of an Erwinia hypersensitive response elicitor protein

PT and related DNA - used to impart disease resistance to plants, to

PT increase their growth and to control insects

XX

XX Claim 4; Page 10-11; 94pp; English.

XX

CC The present sequence represents a hypersensitive response elicitor

CC protein (also called hairpin protein) that is able to elicit a

CC hypersensitive response in plants. The specification also describes

CC hypersensitive response elicitors from other pathogenic organisms.

CC The protein, in non-infectious form, is applied to plants to impart

CC disease resistance (to a wide range of viral, bacterial and fungal

CC pathogens), to improve growth (yield, quantity and quality of seeds,

CC to provide earlier germination etc.) and to control insects (e.g. corn

CC borers, Lepidoptera larvae etc.) The same results are provided by

CC transgenic plants expressing the protein.

XX

XX Sequence 403 AA:

XX SQ

[illegible]

Db 176 gqdgsgssgkqpteqeqnaykkgvtdalsgldnglsqlngg199gsgnagtgl 235
Qy 111 GGNA 114
Db 236 dgss 239

Search completed: June 28, 2002, 10:21:25
Job time: 322 sec

Mon Jul 1 09:28:18 2002

us-09-829-124-2.ra1

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OK protein - protein search, using sw model

Run on: June 28, 2002, 10:16:33 ; Search time 13.3 Seconds
(without alignments)
209.362 Million cell updates/sec

Title: US-09-829-124-2

Perfect score: 584

Sequence: 1 MDSIGNFNSIGNLQPMGIG.....OGGAGMGGGVSSSLAGNA 114

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 04

Maximum Match 1004

Listing first 45 summaries

Database : Issued Patents.AA.*
1: /cgn2_6/ptodata/2/laa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/laa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/laa/5C.COMB.pep.*
4: /cgn2_6/ptodata/2/laa/5D.COMB.pep.*
5: /cgn2_6/ptodata/2/laa/5E.COMB.pep.*
6: /cgn2_6/ptodata/2/laa/5F.COMB.pep.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	111	19.0	344	1	US-08-891-254-7
2	111	19.0	344	2	US-08-819-539-7
3	111	19.0	344	2	US-09-030-270A-7
4	111	19.0	344	4	US-09-984-207-7
5	111	19.0	344	4	US-09-013-587-7
6	111	19.0	344	5	PCT-US96-08819-7
7	111	19.0	385	1	US-08-891-254-3
8	111	19.0	385	2	US-08-819-539-3
9	111	19.0	385	5	PCT-US96-08819-3
10	111	19.0	403	2	US-09-030-270A-3
11	111	19.0	403	2	US-08-851-376A-3
12	111	19.0	403	4	US-08-984-207-3
13	111	19.0	403	4	US-09-013-587-3
14	111	19.0	403	4	US-09-030-270A-10
15	111	19.0	20	2	US-08-984-207-10
16	94	16.1	20	4	US-09-013-587-10
17	94	16.1	20	4	US-09-013-587-10
18	91.5	15.7	424	4	US-08-984-207-10
19	88	15.1	2414	1	PCT-US95-04682-2
20	88	15.1	2414	5	US-08-227-536-2
21	88	15.1	2414	1	US-08-120-927-2
22	88	15.1	2414	4	US-08-851-376A-2
23	77	13.2	739	4	US-08-851-376A-2
24	74.5	12.8	653	4	US-09-061-764A-15
25	74.5	12.8	686	4	US-08-789-333F-41
26	73	12.5	124	4	US-08-789-333F-41
27	73	12.5	173	4	US-08-789-333F-42

28	73	12.5	903	2	US-08-853-310-2
29	72	12.3	369	2	US-08-991-300-2
30	71.5	12.2	485	2	US-08-749-391-2
31	71.5	12.2	485	4	US-09-390-200-2
32	70.5	12.1	760	1	US-08-195-152-2
33	70	12.0	795	1	US-07-716-827C-5
34	67.0	12.0	2703	1	US-08-195-432-19
35	66.5	11.6	599	4	US-09-228-986-117
36	66.5	11.5	481	4	US-09-115-446-2
37	66.5	11.4	482	1	US-08-934-075-1
38	66.5	11.4	482	1	US-08-934-075-1
39	66.5	11.4	482	1	US-08-478-674-1
40	66	11.3	816	2	US-08-785-310A-8
41	66	11.3	816	2	US-08-816-693A-53
42	66	11.3	816	3	US-08-885-291-53
43	66	11.3	816	4	US-09-496-672-53
44	66	11.3	1004	4	US-08-916-352-2
45	65.5	11.2	255	4	US-09-342-084-11

ALIGNMENTS

RESULT 1
US-08-891-254-7
: Sequence 7, Application US/08891254
: Patent No. 5776889
: GENERAL INFORMATION:
: APPLICANT: Wei, Zhong-Min
: TITLE OF INVENTION: Hypersensitive Response
: NUMBER OF SEQUENCES: 9
: CROSS-REFERENCE TO RELATED APPLICATIONS:
: ADDRESS: Hargrave, Devans & Doyle
: STREET: 11000 Square, P.O. Box 1051
: CITY: Rochester
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent In Release #1.0, Version #1.30
: CURRENT APPLICATION DATA: US/08/891,254
: FILING DATE: 10-31-1997
: CLASSIFICATION: 51
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/475,775
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldman, Michael L.
: REGISTRATION NUMBER: 30,727
: REFERENCE/DOCKET NUMBER: 14603/10050
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (716) 263-1304
: TELEFAX: (716) 263-1600
: INFORMATION FOR SEQ ID NO: 7:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 344
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-891-254-7

Query Match 19.0% Score 111; DB 1; Length 344;
Best Local Similarity 25.8%; Pred. No. 7.7e-05;
Matches 40; Conservative 17; Mismatches 54; Indels 44; Gaps 5;

```

QY      3 SIGNPSNIGLQTH-----GIGPOOHEDSSQQSPSAGSE-----OOLOO 42
DB      58 SAGGTGNTGNAPKDNAGANDSKNDPSKSAQPSAKNTGYVDDANNOPHQAQMO 117
QY      43 LLAMFTMMH-----LQOSQSDSANDECG--NEOPONGOGGLSLPTOMLMQIVMLQMN- 94
DB      118 LLEDVLKLKAALHMQQPGNDKGNGVGANGAKAGGGCGLAEQLTEIILAQLOGGG 177
QY      95 -----OGGAMGGGGGSVNSLGN 113
DB      178 ACAGAGGVGGAGGADGGSGACGAGGANGADGEN 212

RESULT    2
          : US-08-819-539-7
          : Sequence 7, Application US/08819539
          : Patent No. 5859324
          : GENERAL INFORMATION:
          : APPLICANT: Wei, Zhong-Min
          : APPLICANT: Beer, Steven V.
          : TITLE OF INVENTION: Hypersensitive Response
          : TITLE OF INVENTION: Induced Resistance In Plants
          : CROSS-REFERENCE TO RELATED APPLICATIONS:
          : PRIORITY CLAIMING ADDRESS:
          : STREET: Clifton Square, P.O. Box 1051
          : CITY: Rochester
          : STATE: New York
          : COUNTRY: U.S.A.
          : ZIP: 14603
          : COMPUTER READABLE FORM:
          : MEDIUM TYPE: Floppy disk
          : COMPUTER: IBM PC compatible
          : OPERATING SYSTEM: PC-DOS/MS-DOS
          : SOFTWARE: Patent In Release #1.0, Version #1.30
          : COPYRIGHT: Copyright © 1988 by US/08/819,539
          : FILING DATE: 17-Mar-1987
          : CLASSIFICATION: 800
          : PRIOR APPLICATION DATA:
          : APPLICATION NUMBER: 08/475,775
          : FILING DATE:
          : ATTORNEY/AGENT INFORMATION:
          : NAME: Goldman, Michael L.
          : REGISTRATION NUMBER: 30,727
          : REFERENCE/DOCKET NUMBER: 14603/10050
          : TELECOMMUNICATION INFORMATION:
          : TELEPHONE: (716) 263-1304
          : TELEFAX: (716) 263-1600
          : INFORMATION FOR SEQ ID NO: 7:
          : SEQUENCE CHARACTERISTICS:
          : LENGTH: 17 amino acids
          : TYPE: amino acid
          : STRANDEDNESS:
          : TOPOLOGY: linear
          : MOLECULE TYPE: protein
          : US-08-819-539-7

Query Match           19.0%   Score 111; DB 2; Length 344;
Best Local Similarity 25.8%; Pred.No. 7.7e-05;
Matches 40; Conservative 17; Mismatches 54; Indels 44; Gaps 5;

QY      3 SIGNPSNIGLQTH-----GIGPOOHEDSSQQSPSAGSE-----OOLOO 42
DB      58 SAGGTGNTGNAPKDNAGANDSKNDPSKSAQPSAKNTGYVDDANNOPHQAQMO 117
QY      43 LLAMFTMMH-----LQOSQSDSANDECG--NEOPONGOGGLSLPTOMLMQIVMLQMN- 94
DB      118 LLEDVLKLKAALHMQQPGNDKGNGVGANGAKAGGGCGLAEQLTEIILAQLOGGG 177
QY      95 -----OGGAMGGGGGSVNSLGN 113

```

```

1 ZIP: 14603
2
3 COMPUTER READABLE FORM:
4
5 MEDIUM TYPE: Floppy disk
6
7 COMPUTER: IBM PC compatible
8
9 OPERATING SYSTEM: PC-DOS/MS-DOS
10
11 SOFTWARE: PatentIn Release #1.0, Version #1.30
12
13 CURRENT APPLICATION DATA:
14
15 APPLICATION NUMBER: US/09/013,587
16
17 FILING DATE:
18
19 CLASSIFICATION:
20
21 PRIOR APPLICATION DATA: US 60/036,048
22
23 APPLICATION NUMBER: 1997
24
25 PRIORITY:
26
27 ATTORNEY/AGENT INFORMATION:
28
29 NAME: Goldman, Michael L.
30
31 REGISTRATION NUMBER: 30,727
32
33 REFERENCE/DOCKET NUMBER: 19603/1501
34
35 TELECOMMUNICATION INFORMATION:
36
37 TELEPHONE: (716) 263-1304
38
39 TELEFAX: (716) 263-1600
40
41 INFORMATION FOR SEQ ID NO: 7:
42
43 SEQUENCE CHARACTERISTICS:
44
45 LENGTH: 344 amino acids
46
47 TYPE: amino acid
48
49 STRANDS: 1
50
51 TOPOLOGY: linear
52
53 MOLECULE TYPE: protein
54
55 US-09-013-587-7

```

118	LLEDLVLLKKAALHMOQPGGNDKNGCVGGANGARGAGGGGLAENLQETQILAQAGGG	177
119	-----	
120	-----	
121	-----	
122	-----	
123	-----	
124	-----	
125	-----	
126	-----	
127	-----	
128	-----	
129	-----	
130	-----	
131	-----	
132	-----	
133	-----	
134	-----	
135	-----	
136	-----	
137	-----	
138	-----	
139	-----	
140	-----	
141	-----	
142	-----	
143	-----	
144	-----	
145	-----	
146	-----	
147	-----	
148	-----	
149	-----	
150	-----	
151	-----	
152	-----	
153	-----	
154	-----	
155	-----	
156	-----	
157	-----	
158	-----	
159	-----	
160	-----	
161	-----	
162	-----	
163	-----	
164	-----	
165	-----	
166	-----	
167	-----	
168	-----	
169	-----	
170	-----	
171	-----	
172	-----	
173	-----	
174	-----	
175	-----	
176	-----	
177	-----	
178	AGAGAGGGVGGGAGDGGGAGGAGGANGADGNG	212
179	-----	
180	-----	
181	-----	
182	-----	
183	-----	
184	-----	
185	-----	
186	-----	
187	-----	
188	-----	
189	-----	
190	-----	
191	-----	
192	-----	
193	-----	
194	-----	
195	-----	
196	-----	
197	-----	
198	-----	
199	-----	
200	-----	
201	-----	
202	-----	
203	-----	
204	-----	
205	-----	
206	-----	
207	-----	
208	-----	
209	-----	
210	-----	
211	-----	
212	-----	
213	-----	
214	-----	
215	-----	
216	-----	
217	-----	
218	-----	
219	-----	
220	-----	
221	-----	
222	-----	
223	-----	
224	-----	
225	-----	
226	-----	
227	-----	
228	-----	
229	-----	
230	-----	
231	-----	
232	-----	
233	-----	
234	-----	
235	-----	
236	-----	
237	-----	
238	-----	
239	-----	
240	-----	
241	-----	
242	-----	
243	-----	
244	-----	
245	-----	
246	-----	
247	-----	
248	-----	
249	-----	
250	-----	
251	-----	
252	-----	
253	-----	
254	-----	
255	-----	
256	-----	
257	-----	
258	-----	
259	-----	
260	-----	
261	-----	

RESULT 6
CT-US96-08819-7
Sequence 7, Application PC/TU9608819
GENERAL INFORMATION:
APPLICANT: Cornell Research Foundation, Inc.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED
TITLE OF INVENTION: RESISTANCE IN PLANTS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
Nixon, Hargrave, Devans & Doyle LLP
ATTORNEYS
STREET: CitiBank Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE PATENT IN RELEASE #1.0, Version #1.30
CURSABLE AND DATA FILED
APPLICATION NUMBER: PCT/US96/08819
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
PRIOR APPLICATION NUMBER: US 08/475,775

```

: FILING DATE: 07-JUN-1995
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldman, Michael L.
: REGISTRATION NUMBER: 30,727
: REFERENCE/DOCKET NUMBER: 14603/10051
: TELEPHONE: (716) 263-1304
: TELEFAX: (716) 263-1304
: INFORMATION FOR SEQ ID NO: 7:
: LENGTH: 344 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
PCT-US96-08619-7

Query Match 19.04; Score 111; DB 5; Length 344;
Best Local Similarity 25.88; Pred. No. 7.9e-05;
Matches 40; Conservative 17; Mismatches 54; Indels 44; Gaps 5;

Oy 3 SIGNNFSNIGNLQTM-----GIGPOHEDSSQSPSAGSE-----QQLDQ 42
Db 58 SAGGTCNTGNAFAGKGNAGNAGNDPSKNDPSKQASQASANKTGNVDANNDPQALMQ 117
Oy 43 LLMFIMM-----LQSSGSDANOEG--NEQPONGQSGLSPLTOMLQIYVMOLMKN- 94
Db 118 LLEDVLLKLLALHMQPGNGKNGVGGANGKAGGQGGGLAEALQETDILQALGGG 177
Oy 95 -----QGGAGGGGGSVNSSLGN 113
Db 178 AGAGAGGGVGGAGGAGGSGGAGGANGAGADGN 212

RESULT 7
US-08-891-254-3
: Sequence 3, Application US/08891254
: Patent No. 5776889
: GENERAL INFORMATION:
: APPLICANT: Wei, Zhong-Min
: APPLICANT: Beer, Steven V.
: TITLE OF INVENTION: Hypersensitive Response
: TITLE OF INVENTION: Induced Resistance in Plants
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon Hatgrave, Devans & Doyle
: CITY: Rochester
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent 1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/891,254
: FILING DATE: 10-JUL-1997
: CLASSIFICATION: 514
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/475,775
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldman, Michael L.
: REGISTRATION NUMBER: 30,727
: REFERENCE/DOCKET NUMBER: 14603/10050
: TELEPHONE: (716) 263-1304
: TELEFAX: (716) 263-1304
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 385 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-819-539-3

Query Match 19.04; Score 111; DB 2; Length 385;
Best Local Similarity 31.54; Pred. No. 8.9e-05;
Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;

Oy 5 GNNFSNIGNL--OTMGI-GPOHEDSSQSPS-AGSPQOLQILAMFIMMQLQSS--O 56
Db 118 GNNFTTSTNPLQALGINTSNDSTGTSTSDSSDPMOOLLKHFSEIM--QSLFGD 175
Oy 57 GSDANO--ECNGNEPONGQO-----EGLSPLTOMLQIYVMOLMKNQGGAGGGGVNSSL 110
Db 176 GDDGTQSSSSGKQPTGEONAYKGYDALSGLMGLNGLSOLLGNGGLGGGGGNAAGTGL 235
Oy 111 GGA 114
Db 236 DGSS 239

RESULT 8
US-08-819-539-3
: Sequence 3, Application US/08819539
: Patent No. 5859324
: GENERAL INFORMATION:
: APPLICANT: Wei, Zhong-Min
: APPLICANT: Beer, Steven V.
: TITLE OF INVENTION: Hypersensitive Response
: TITLE OF INVENTION: Induced Resistance in Plants
: NUMBER OF SEQUENCES: 9
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Nixon Hatgrave, Devans & Doyle
: CITY: Rochester
: STATE: New York
: COUNTRY: U.S.A.
: ZIP: 14603
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent 1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/819,539
: FILING DATE: 17-MAR-1997
: CLASSIFICATION: 800
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/475,775
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Goldman, Michael L.
: REGISTRATION NUMBER: 30,727
: REFERENCE/DOCKET NUMBER: 14603/10050
: TELEPHONE: (716) 263-1304
: TELEFAX: (716) 263-1304
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 385 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: linear
: MOLECULE TYPE: protein
US-08-819-539-3
```

Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;

OY 5 GNNFSNIGNL---QTMGI-GPOOHEDSSQSPS-AGSEQQLDQLLAFIMHMLAQSS---Q 56
DB 118 GNNFTSTNSPLDQALGINSTSDSTSDSDPQQLKMFSEIM--OSLFQD 175
OY 57 GSDANQ--ECGNEOPONGQ---BELSPLTOMLQIWMQMGAGGAGGGSVNSL 110
DB 176 GQDGTQSSSGKOPTGEGONAYKGVTDALSLGKNGLSQLLNGGLGGGAGNACTGL 235
OY 111 GGNA 114
DB 236 DGSS 239

RESULT 9

PCT-US93-06243-2
SEQUENCE 2, Application PC/TUS9306243
GENERAL INFORMATION:
APPLICANT: Zhong-Min Wei, David W. Bauer, Steven V.
APPLICANT: Beer, Alan Collier, Sheng-Yang He, and Ron J. Laby.
TITLE OF INVENTION: Elicitor of the Hypersensitive Response in Plants
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: Yahwak & Associates
STREET: 25 Skytop Drive
CITY: Trumbull
STATE: Connecticut
COUNTRY: U.S.A.
ZIP: 06611
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: Macintosh
OPERATING SYSTEM: MS-DOS
SOFTWARE: Microsoft Word 4.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/06243
FILING DATE: 19930630
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 907,935
FILING DATE: 01 JUL 1992
ATTORNEY/AGENT INFORMATION:
NAME: George M. Yahwak
REGISTRATION NUMBER: 26,824
REFERENCE/DOCKET NUMBER: CRF D-1172
TELECOMMUNICATION INFORMATION:
TELEPHONE: (203)268-1951
TELEFAX: (203)268-1951
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 385 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
PCT-US93-06243-2

Query Match 19.0%; Score 111; DB 5; Length 385;
Best Local Similarity 31.5%; Pred. No. 8.9e-05;
Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;

OY 5 GNNFSNIGNL---QTMGI-GPOOHEDSSQSPS-AGSEQQLDQLLAFIMHMLAQSS---Q 56
DB 118 GNNFTSTNSPLDQALGINSTSDSTSDSDPQQLKMFSEIM--OSLFQD 175
OY 57 GSDANQ--ECGNEOPONGQ---BELSPLTOMLQIWMQMGAGGAGGGSVNSL 110
DB 176 GQDGTQSSSGKOPTGEGONAYKGVTDALSLGKNGLSQLLNGGLGGGAGNACTGL 235
OY 111 GGNA 114
DB 236 DGSS 239

RESULT 11

US-08-200-7244-2
SEQUENCE 2, Application US/082007244
PATENT NO. 5849868
GENERAL INFORMATION:
APPLICANT: Wei, Zhong-Min
APPLICANT: Bauer, David W.
APPLICANT: Beer, Steven V.
APPLICANT: Collier, Alan

DB 236 DGSS 239

RESULT 10

PCT-US96-08819-3
SEQUENCE 3, Application PC/TUS9608819
GENERAL INFORMATION:
APPLICANT: Cornell Research Foundation, Inc.
TITLE OF INVENTION: HYPERSENSITIVE RESPONSE INDUCED BY RESISTANCE IN PLANTS
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
STREET: Clinton Square, P.O. Box 1051
CITY: Rochester
STATE: New York
COUNTRY: U.S.A.
ZIP: 14603
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/08819
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/475,775
FILING DATE: 07 JUN 1995
ATTORNEY/AGENT INFORMATION:
NAME: Goldman, Michael
REGISTRATION NUMBER: 10,727
REFERENCE/DOCKET NUMBER: 19603/10051
TELECOMMUNICATION INFORMATION:
TELEPHONE: (716) 263-1304
TELEFAX: (716) 263-1600
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 385 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
MOLECULE TYPE: protein
PCT-US96-08819-3

Query Match 19.0%; Score 111; DB 5; Length 385;
Best Local Similarity 31.5%; Pred. No. 8.9e-05;
Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;

OY 5 GNNFSNIGNL---QTMGI-GPOOHEDSSQSPS-AGSEQQLDQLLAFIMHMLAQSS---Q 56
DB 118 GNNFTSTNSPLDQALGINSTSDSTSDSDPQQLKMFSEIM--OSLFQD 175
OY 57 GSDANQ--ECGNEOPONGQ---BELSPLTOMLQIWMQMGAGGAGGGSVNSL 110
DB 176 GQDGTQSSSGKOPTGEGONAYKGVTDALSLGKNGLSQLLNGGLGGGAGNACTGL 235
OY 111 GGNA 114
DB 236 DGSS 239

APPLICANT: He, Sheng-Yang
 APPLICANT: Laby, Ron J.
 TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE
 IN PLANTS
 NUMBER OF SEQUENCES: 5
 CORRESPONDENCE ADDRESS:
 ADDRESS: Milton Margrave, Devans & Doyle
 STREET: Clinton Square
 CITY: Rochester
 STATE: New York
 COUNTRY: U.S.A.
 ZIP: 14603
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/200,724A
 FILING DATE: 23-FEB-1994
 CLASSIFICATION: 530
 ATTORNEY/AGENT INFORMATION:
 NAME: Goldman, Michael L.
 REGISTRATION NUMBER: 30,727
 REFERENCE/DOCKET NUMBER: 19603/10030
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304
 TELEFAX: (716) 263-1600
 INFORMATION FOR SEQ ID NO: 2:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 403 amino acids
 TYPE: amino acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: peptide
 US-08-200-724A-2

Query Match 19.04; Score 111; DB 2; Length 403;
 Best Local Similarity 31.54; Pred. No. 9.4e-05;
 Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;
 Oy 5 GNFSNIGNL---OTMGI-GROQHESSQSPS-ACSEQQLQQLLAFIMHMQQS---Q 56
 Db 118 GNNITSTNPLDQALGINSTSDSTSGTSDSDPMQOLLKMFSEIM--QSLPFD 175
 Oy 57 GSDANO--ECGNBPQNGOQ---EGLSPLTMQIMQIMQNGAGCGGGSVNSL 110
 Db 176 GQDTGCGSSSGKQPTGEONAYKGVTDALSGLMGSLGSLGAGGCGGNGTGL 235
 Oy 111 GQNA 114
 Db 236 DGSS 239

RESULT 12
 US-09-030-270A-3
 : Sequence 3, Application US/09030270A
 : Patent No. 5977060
 : GENERAL INFORMATION:
 : APPLICANT: Zitter, Thomas A.
 : APPLICANT: Wei, Zhong-Min
 : APPLICANT: Collmer, Alan
 : APPLICANT: He, Sheng-Yang
 : APPLICANT: Laby, Ron
 : TITLE OF INVENTION: INSECT CONTROL WITH A
 : HYPERSENSITIVE RESPONSE ELICITOR
 : NUMBER OF SEQUENCES: 10
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Milton Margrave, Devans & Doyle LLP
 : STREET: P.O. Box 1051, Clinton Square
 : CITY: Rochester
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 14603
 : COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/09/030,270A
 FILING DATE:
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 60/039,226
 FILING DATE: 28-FEB-1997
 ATTORNEY/AGENT INFORMATION:
 NAME: Goldman, Michael L.
 REGISTRATION NUMBER: 30,727
 REFERENCE/DOCKET NUMBER: 19603/1521
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (716) 263-1304
 TELEFAX: (716) 263-1600
 INFORMATION FOR SEQ ID NO: 3:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 403 amino acids
 TYPE: amino acid
 STRANDEDNESS:
 TOPOLOGY: linear
 MOLECULE TYPE: Protein
 US-09-030-270A-3

Query Match 19.04; Score 111; DB 2; Length 403;
 Best Local Similarity 31.54; Pred. No. 9.4e-05;
 Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;
 Oy 5 GNFSNIGNL---OTMGI-GROQHESSQSPS-ACSEQQLQQLLAFIMHMQQS---Q 56
 Db 118 GNNITSTNPLDQALGINSTSDSTSGTSDSDPMQOLLKMFSEIM--QSLPFD 175
 Oy 57 GSDANO--ECGNBPQNGOQ---EGLSPLTMQIMQIMQNGAGCGGGSVNSL 110
 Db 176 GQDTGCGSSSGKQPTGEONAYKGVTDALSGLMGSLGSLGAGGCGGNGTGL 235
 Oy 111 GQNA 114
 Db 236 DGSS 239

RESULT 13
 US-08-851-376A-2
 : Sequence 2, Application US/08851376A
 : Patent No. 6347170
 : GENERAL INFORMATION:
 : APPLICANT: Beer, Steven V.
 : APPLICANT: Wei, Zhong-Min
 : APPLICANT: Bauer, David W.
 : APPLICANT: Collmer, Alan
 : APPLICANT: He, Sheng-Yang
 : APPLICANT: Laby, Ron
 : TITLE OF INVENTION: ELICITOR OF THE HYPERSENSITIVE RESPONSE
 : IN PLANTS
 : NUMBER OF SEQUENCES: 5
 : CORRESPONDENCE ADDRESS:
 : ADDRESS: Maxon Peabody LLP
 : STREET: Clinton Square, P.O. Box 1051
 : CITY: Rochester
 : STATE: NY
 : COUNTRY: U.S.A.
 : ZIP: 14603
 : COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: Patentin Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/851,376A

US-09-013-587-3

Query Match 19.0%; Score 111; DB 4; Length 403;
Best Local Similarity 31.5%; Pred. No. 9.4e-05;
Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;
OY 5 GNVFSNIGNL---QTMGI-GPOHEDSSQSPS-AGSEBQIDQLLAFIMMILQOS---Q 56
DB 118 GNVYTSNPLDQALGINSTSQNDSTSTSTSDSPHQQLLAFSEIM--QSLFGD 175
OY 57 GSDANO--ECGNEQFONGQ-----EGLSPLTOMLQITVQLMONGGAGMCGGGSVNSL 110
DB 176 GQDGTGSSGCGKQPTGEBQNNYKGVTDALSGLMNGLSQLLGNGLGCGGCGNAGTGL 235
OY 111 GGA 114
DB 236 DGS 239

Search completed: June 28, 2002, 10:21:52
Job time: 319 sec

Query Match 19.0%; Score 111; DB 2; Length 403;
Best Local Similarity 31.3%; Pred. No. 0.0019;
Matches 39; Conservative 21; Mismatches 48; Indels 16; Gaps 7;
QY 5 GNENFIGNL--QTGMI-GPQHEDSSQSPS-AGSEQDQLDQLAMFMMMLQOQS---Q 56
DB 118 GNTTSTTSPDQALGINSSTDSSTSTSTSDSDPMQQLKMFESIM--OSLFGD 175
QY 57 GSDAMQ--ECGNEQDPONQ---BELSPITOMLQVYVMOIMQNGGAGMGGSVNSLS 110
DB 176 QDGTGSSGCKGQPTGEGNQYKATGVTGDSLGLNGLSQLLGNGLGSGGGGAGTGL 235
QY 111 GGVA 114
DB 236 DGSS 239
RESULT 3
C75348
Hypothetical protein - Deinococcus radiodurans (strain R1)
C:Species: Deinococcus radiodurans
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 31-Mar-2000
R:Accession: C75348 J.A.; Heidelberg, J.F.; Hickey, E.K.; Peterson, J.D.; Dodson, R.J.;
M.; Shen, M.; Vamathevan, J.; Adams, P.; McDonald, L.; Utterback, T.; Zalewski, C.; Ma
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioreistant bacterium Deinococcus radiodurans R1.
A:Reference number: A75250; MUID:20036896
A:Accession: C75348
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-272 <WHI>
A:Cross-references: GB:AE001889; GB:AE000513; NID:96457865; PIDN:AAF09792.1; PID:9645787
A:Experimental source: strain R1
A:GenBank: C75348
A:Map position: 1
Query Match 16.5%; Score 96.5; DB 2; Length 272;
Best Local Similarity 25.5%; Pred. No. 0.033;
Matches 39; Conservative 15; Mismatches 36; Indels 63; Gaps 7;
QY 1 MDSIGNFENIGNLQTMGIPGQHESSQSPSAGSEQDQLDQLAMFMMMLQOQSQSDA 60
DB 41 LGAMTRATPGGLDALASAVGDDSA-----MLGFGQGTGDM 80
QY 61 NQ-----ECNEQDPON---GQEGSLPTQMLQIVM-----Q 90
DB 81 NEQKILGHVFGQDAAANNIGDIPD--QNMGLIYAVAPILSLYLSNRSRAQGG 138
QY 91 LMQNGG-----AGMG--GGSVNSLGG 112
DB 139 MCGNNGGSGGMSLPGNGGGLGGLGSLGG 171
RESULT 4
A33106
neurogenic locus mmm protein - fruit fly (Drosophila melanogaster)
A:Alternative names: mastermind protein
A:Title: Neurogenic locus mmm protein
C:Date: 07-Sep-1990 #sequence_revision 07-Sep-1990 #text_change 21-Jul-2000
R:Accession: A33106 J.A.; Schaller, D.; Friedel, C.; Schmid, A.; Bettler, D.; Lam, L.; Yedvobnick, B.
Genes Dev. 4, 1688-1700, 1990
A:Title: The Drosophila neurogenic locus mastermind encodes a nuclear protein unusually
A:Reference number: A36391; MUID:91065516
A:Accession: A36391
A:Status: preliminary
A:Molecule type: mRNA
A:Residues: 1-1596 <SHO>

A:Cross-references: GB:M54251; NID:98203; PIDN:CAA38152.1; PID:98204
A:Note: strain Canton S
C:Genetics:
A:Genes: FlyBase:mam
A:Cross-references: FlyBase:FBgn0002643

Query Match 16.5%; Score 96.5; DB 2; Length 1596;
Best Local Similarity 28.8%; Pred. No. 0.22;
Matches 32; Conservative 7; Mismatches 39; Indels 33; Gaps 4;
QY 22 QHREDSQSPSAGSEQDQLDQLAMFMMMLQOQSQSDANQCGNEQDPONQOQEGSLPT 81
DB 1428 QQQQQQQHMGFGAANNMOQLLOO-----QSSGGG-----GNMAMOMTSMH-MT 1474
QY 82 QMLMQIVQMLN-----QGGAGMGGSVNSLGG 112
DB 1475 QTOQQITMQQQQPVGVSTTTTHQOQOQWQMGPGGGGGGGGPGCSANNNGG 1525

RESULT 5
G86385
Hypothetical protein F237.4 [Imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
R:Accession: G86385
A:Title: G86385
A:Authors: J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alu
A:Reference number: A86141; MUID:21016719
A:Accession: G86385
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-538 <STO>
A:Cross-references: GB:AE005172; NID:910092326; PIDN:AGI2737.1; GSPDB:GN00141
A:Genetics:
A:Map position: 1

Query Match 16.4%; Score 96; DB 2; Length 534;
Best Local Similarity 34.1%; Pred. No. 0.077;
Matches 31; Conservative 4; Mismatches 44; Indels 12; Gaps 2;
QY 22 QHREDSQSPSAGSEQDQLDQLAMFMMMLQOQSQSDANQCGNEQDPONQOQEGSLPT 81
DB 436 QHQQ 491
QY 82 QMLMQIVQMLN-----QGGAGMGGSVNSLGG 104
DB 492 QNQOQQQPOQWNGGQAFAPGSGGQGGG 522

RESULT 6
T22330
Hypothetical protein F47A4.2 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C:Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 18-Feb-2000
R:Accession: T22330
A:Title: T22330
A:Authors: B.
Submitted to the EMBL Data Library, June 1995
A:Reference number: A21545
A:Accession: T22330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-3498 <WIL>
A:Cross-references: EMBL:249888; PIDN:CAA90064.1; GSPDB:GN00028; CESP:F47A4.2

A: Experimental source: clone F47A4

C: Gene: CESP:F47A4.2

A: Map position: X

A: Introns: 36/3; 72/3; 147/2; 228/3; 261/2; 320/2; 847/3; 881/1; 1317/2; 1500/3; 1653/2;

Query Match

Best Local Similarity 15.84; Score 92.5; DB 2; Length 3498;

Matches 32; Conservative 4; Mismatches 42; Indels 7; Gaps 4;

OY 20 GPOOHEDSSQSPSAGSEQQDQLLAFIMMLQSQSGSDANQECNQPQNGQOGLSP 79

DB 3369 GCOOHQ--PQOQSIQOQOQOQDQTRMOAOMQOQPTAQQOQNRMG--MPSQ-QQSGNAY 3423

OY 80 LTOHLMQVIMQMLQNGGAGGGG 104

DB 3424 SNMQFGVHQGGG--GNGGSG 3446

RESULT 7

H85294

Hypothetical protein Atg25520 [imported] - Arabidopsis thaliana

C: Species: Arabidopsis thaliana (mouse-ear cress)

C: Date: 16-Feb-2001 #sequence_revision 16-Feb-2001 #text_change 16-Feb-2001

C: Accession: H85294

C: Title: The European Union Arabidopsis Genome Sequencing Consortium, The Cold Spring Harbor Laboratory, 1999

A: Title: Sequence and analysis of chromosome 4 of the plant Arabidopsis thaliana.

A: Reference number: H85001; MUID:2003488

A: Accession: H85294

A: Status: preliminary

A: Molecule type: DNA

A: Residues: 1-748 <STO>

A: Cross-references: GB:NC_001268; MID:97269402; PIDN:CAB81362.1; GSPDB:GN00140

C: Gene: Atg25520

A: Map position: 4

Query Match

Best Local Similarity 15.58; Score 90.5; DB 2; Length 748;

Matches 37; Conservative 18; Mismatches 55; Indels 27; Gaps 5;

OY 2 DSGNFSNIGNLQTMGTPQOHHDSQSPSAGSEQQDQLLAFIMMLQSQSGSDAN 61

DB 572 NAMNPNSTGKQF--GFSSQNTFNSNSPSSSQOQRHLVTGGFPNQOQRTMN 629

OY 62 QECNEQPONGQOE-----GLSPLTQMLQVIMQMLQNGG-----AGMGGGGS-- 105

DB 630 GPT-NILPQNHQLOSPHSHGNTPEOQMLHQLQEMSENGSVQOQQAFAFGSGSGNSNA 688

OY 106 -----VNSSLGNA 114

DB 689 ERMTTASTNISGGRA 705

RESULT 8

T05795

Hypothetical protein M732.110 - Arabidopsis thaliana

C: Species: Arabidopsis thaliana (mouse-ear cress)

C: Date: 09-Apr-1999 #sequence_revision 09-Apr-1999 #text_change 13-Aug-1999

C: Accession: T05795

C: Title: The Arabidopsis Genome Sequencing Consortium, April 1998

A: Reference number: T13492

A: Accession: T05795

A: Status: DNA

A: Residues: 1-748 <BEV>

A: Cross-references: EMBL:AL022197

A: Experimental source: cultivar Columbia; BAC clone M732

C: Genetics:

A: Map position: 4

A: Introns: 205/3; 249/3; 266/1; 337/3; 360/3; 394/2; 425/3; 449/1

A: Note: M7J2.110

Query Match

Best Local Similarity 15.58; Score 90.5; DB 2; Length 748;

Matches 37; Conservative 18; Mismatches 55; Indels 27; Gaps 5;

OY 2 DSGNFSNIGNLQTMGTPQOHHDSQSPSAGSEQQDQLLAFIMMLQSQSGSDAN 61

DB 572 NAMNPNSTGKQF--GFSSQNTFNSNSPSSSQOQRHLVTGGFPNQOQRTMN 629

OY 62 QECNEQPONGQOE-----GLSPLTQMLQVIMQMLQNGG-----AGMGGGGS-- 105

DB 630 GPT-NILPQNHQLOSPHSHGNTPEOQMLHQLQEMSENGSVQOQQAFAFGSGSGNSNA 688

OY 106 -----VNSSLGNA 114

DB 689 ERMTTASTNISGGRA 705

RESULT 9

T26216

Hypothetical protein W06B11.2 - Caenorhabditis elegans

C: Species: Caenorhabditis elegans

C: Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999

C: Accession: T26216

R: Fulton, L.

A: Submitted to the EMBL Data Library, November 1995

A: Description: The sequence of C. elegans cosmid W06B11.

A: Reference number: Z20174

A: Accession: T26216

A: Status: Preliminary

A: Molecule type: DNA

A: Residues: 1-707 <EUL>

A: Cross-references: EMBL:U39854; PIDN:AA81077.1; CESP:W06B11.2

C: Genetics:

A: Gene: CESP:W06B11.2

A: Introns: 17/1; 72/3; 88/2; 232/1; 397/3; 494/3; 606/2; 665/1

Query Match

Best Local Similarity 15.44; Score 90; DB 2; Length 707;

Matches 36; Conservative 7; Mismatches 21; Indels 72; Gaps 6;

OY 8 FSGTGNLQF-----WGIGPQO-----HED-----SSQGS 31

DB 197 YQNNMLTTFSGANGLPQOIQORPQOQFPIHEPLOGHAGNGFGQOYFFQKQNGH 256

OY 32 PSAGSEQQDQLLAFIMMLQSQSGSDANQECNQPQNGQOGLSPITQMLQVIMOL 91

DB 257 PQGQAQQQLQQL-----AQQHQ-----QQQNSQO----- 280

RESULT 10

T13828

CREB-binding protein homolog - fruit fly (Drosophila melanogaster)

C: Species: Drosophila melanogaster

C: Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 17-Nov-2000

C: Accession: T13828

R: Akamaru, H.; Chen, Y.; Dai, P.; Hou, D.X.; Nonaka, M.; Smolik, S.M.; Armstrong, S.

A: Title: Drosophila CBP is a co-activator of cubitus interruptus in hedgehog signal

A: Reference number: T17785; MUID:97263578

A: Accession: T13828

A: Status: Preliminary

A: Molecule type: mRNA

A: Residues: 1-390 <AK>

A: Cross-references: EMBL:U08570; MID:g1916929; PID:g1916930; PIDN:AA853050.1

[illegible]

CHANG, J. Y. 1993. *Journal of the American Water Resources Association* 29:103-114.

Search completed: June 28, 2002, 10:22:21
Job time: 323 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:20:53 ; Search time 22.09 Seconds
(Without alignments)
892.776 Million cell updates/sec

Title: US-09-829-124-2
Perfect score: 584
Sequence: 1 MDSIGNNFSNIGLQTMGIG.....OCGAGMGGSVNSLAGNA 114

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

- Database : SPTREMBL19.*
- 1: sp.archaea.*
 - 2: sp.bacteria.*
 - 3: sp.fungi.*
 - 4: sp.human.*
 - 5: sp.invertebrate.*
 - 6: sp.mammal.*
 - 7: sp.mhc.*
 - 8: sp.organelle.*
 - 9: sp.phage.*
 - 10: sp.plant.*
 - 11: sp.protist.*
 - 12: sp.virus.*
 - 13: sp.vertebrate.*
 - 14: sp.unclassified.*
 - 15: sp.virus.*
 - 16: sp.bacteriap.*
 - 17: sp.archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	189.5	32.4	139	2 O9KW32	O9KW32 xanthomonas
2	189	32.4	143	2 O9Z1J5	O9Z1J5 xanthomonas
3	114	19.5	330	2 O9WKK0	O9WKK0 burkholderi
4	110.5	18.9	838	5 O9YOC9	O9YOC9 dictyostell
5	110	18.8	344	2 O9RBS0	O9RBS0 burkholderi
6	104.6	18.2	318	5 O9WOP3	O9WOP3 drosophila
7	104.6	17.2	1173	2 O9RBT0	O9RBT0 burkholderi
8	100.5	17.2	1173	2 O9RBT0	O9RBT0 burkholderi
9	96.5	16.5	272	16 O9RBY5	O9RBY5 deinooccus
10	96.5	16.5	1366	5 O9V6R5	O9V6R5 drosophila
11	96.5	16.5	1594	5 O9V6W7	O9V6W7 drosophila
12	96	16.4	534	10 O9CSM3	O9CSM3 caenorhabdi
13	92.5	15.8	3498	5 Q20497	Q20497 caenorhabdi
14	91.5	15.7	424	2 O87264	O87264 pseudomonas
15	91	15.6	624	5 O94486	O94486 dictyostell
16	90.5	15.5	748	10 O65609	O65609 arabidopsis

17	90.5	15.5	748	10 O9MOK8	O9MOK8 arabidopsis
18	90	15.4	707	5 Q23191	Q23191 caenorhabdi
19	90	15.4	748	5 O95TP4	O95TP4 drosophila
20	90	15.4	2328	5 O9VM47	O9VM47 drosophila
21	90	15.4	2531	5 O9GPH4	O9GPH4 drosophila
22	90	15.4	3190	5 O01368	O01368 drosophila
23	90	15.4	3275	5 O9M321	O9M321 drosophila
24	88	15.2	1783	3 P79025	P79025 emericeila
25	88	15.2	1783	3 P79025	P79025 emericeila
26	87	14.9	1778	5 O9V6R2	O9V6R2 drosophila
27	87	14.9	2035	5 O9G919	O9G919 drosophila
28	86	14.7	794	10 P93560	P93560 pseudomonas
29	85.5	14.6	195	16 O9HXK1	O9HXK1 pseudomonas
30	85.5	14.6	642	3 O9UVW0	O9UVW0 rhizopus st
31	85.5	14.6	2994	5 O95ZG5	O95ZG5 dictyostell
32	85	14.6	147	5 O9V543	O9V543 drosophila
33	85	14.6	365	2 O9KH45	O9KH45 erwinia her
34	85	14.6	713	10 O9SDM2	O9SDM2 triticum ae
35	85	14.6	713	10 O94IK8	O94IK8 secale cere
36	85	14.6	713	10 O94IK8	O94IK8 secale cere
37	85	14.6	713	10 O94IK8	O94IK8 secale cere
38	84.5	14.5	707	5 O9VJZ8	O9VJZ8 secale cere
39	84.5	14.5	915	5 O24433	O24433 drosophila
40	84.5	14.5	915	5 O24433	O24433 drosophila
41	84.5	14.5	915	5 O9VCN7	O9VCN7 drosophila
42	84.5	14.5	2653	5 O25253	O25253 lucilia cup
43	84	14.4	447	2 O54508	O54508 erwinia amy
44	84	14.4	447	2 O9LAW2	O9LAW2 erwinia amy
45	84	14.4	926	5 O9W3G1	O9W3G1 drosophila

ALIGNMENTS

RESULT 1

ID O9KW32 PRELIMINARY: PRT: 139 AA.

AC O9KW32

DT 01-OCT-2000 (TREMREL. 15, Created)

DT 01-OCT-2000 (TREMREL. 15, Last sequence update)

DT 01-DEC-2001 (TREMREL. 19, Last annotation update)

DE HPAL HOMOLOG

OS Xanthomonas oryzae pv. oryzae.

OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;

OX Xanthomonas

OX NEB_L381D-64187;

OX SEQUENCE FROM N.A.

RC STRAIN-MAFF 311018;

RX MEDLINE-21303248; PubMed-11410350;

RA Ochiai H., Inoue Y., Hasebe A., Kaku H.;

RT "Construction and characterization of a Xanthomonas oryzae pv. oryzae

RT Bacterial artificial chromosome library.";

RL FEMS Microbiol. Lett. 200:59-65(2001).

DR EMBL: AB045311; BAB07848.1; -

SQ SEQUENCE 139 AA: 13727 MW: D05857FD8E403898 CRC64;

Query Match 32.4%; Score 189.5; DB 2; Length 139;

Best Local Similarity 40.7%; Pred. No. 3.5e-13;

Matches 50; Conservative 12; Mismatches 48; Indels 13; Gaps 5;

QY 1 MDSIGNNFSNIGLQTMGIGPOQED--SSQSQSPSAG--SSQQLDQLLAFMMMLAQOS---55

DB 1 MNSLNTQFG--GSTSNLQVGFSDPTFGSNGOGISSEKQLDQLLQSLALQSSKNA 58

QY 56 -----OGSDANQECNEDQPNQDQBSLPTQNLQIYVQLMONGAGMGGSVNSL 110

DB 59 BEKGQGGGNGGCGGNSQQAGQONGGSPFTQNLMLHIVGISTQNGGAGGGG--FGG 117

QY 111 GGN 113

DB 118 GGD 120

```

RESULT 2
ID Q921J5 PRELIMINARY; PRT: 143 AA.
AC Q921J5
DT 01-MAY-1999 (TRENBLrel. 10, Created)
DT 01-MAY-1999 (TRENBLrel. 10, Last sequence update)
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)
GN HPA1
OS Xanthomonas oryzae pv. oryzae.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OX NCBI_TaxID=64187;
RN [1]
RP STRAIN=PX086;
RX MEDLINE=20179737; PubMed=10714988;
RA Zhu W., MacBanua M.M., White F.F.;
RT "Identification of two novel hrp-associated genes in the hrp gene
cluster of Xanthomonas oryzae pv. oryzae."
DB EMBL/AB026127; AAC54512.1;
SQ SEQUENCE 143 AA; 31987 MW; 0538C594CA241E5C CRC64;

Query Match 32.4%; Score 189; DB 2; Length 143;
Best Local Similarity 41.0%; Pred. No. 4.1e-13;
Matches 50; Conservative 11; Mismatches 49; Indels 12; Gaps 4;

QY 1 MDISGNFSGNIGLQTMGTGPGQHEH-SSQSPFAG-SEQQLDQLLAFIMMLQAS--- 55
Db 1 WNSLNTQFG-GSTNLQVSPDPTTFSGQGGNGGISEKQLDQLLQALQSSKNA 58
QY 56 -----QGSNDANFQNDPONGQSGLSPLTOMLQIVMLQMGQGGGSSNSL 110
Db 59 ECKGKGDDGGGGGNSQAGQAGQNGPPTUHLHIVEILQNGGGAGGGGFGGFG 118
QY 111 GG 112
Db 119 GG 120

RESULT 3
ID Q9WKKO PRELIMINARY; PRT: 330 AA.
AC Q9WKKO
DT 01-NOV-1999 (TRENBLrel. 12, Created)
DT 01-NOV-1999 (TRENBLrel. 12, Last sequence update)
DT 01-NOV-1999 (TRENBLrel. 12, Last annotation update)
DE POPAI PROTEIN.
GN POPAI.
OS Burkholderia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OX NCBI_TaxID=305;
RN [1]
RP "PopAI ORF of Ralstonia solanacearum."
RT Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL/AB026239; BAA7270.1;
SQ SEQUENCE 330 AA; 31997 MW; 8EC4C2D9E319DA5 CRC64;

Query Match 19.5%; Score 114; DB 2; Length 330;
Best Local Similarity 28.4%; Pred. No. 0.00018;
Matches 40; Conservative 15; Mismatches 56; Indels 30; Gaps 5;

QY 3 SIGNWFSGNL-----QTMGICPQHR-----RDSQSQSPFAGSEQQLDQ 42
Db 111

```

Query Match 18.8%; Score 110; DB 2; Length 344;
Best Local Similarity 25.8%; Pred. No. 0.0005;
Matches 40; Conservative 16; Mismatches 55; Indels 44; Gaps 5

QY	3	STGNFNSNIGLQIOTN	-----GIGPOHDSQSQSPAGS	-----QOLDQ 43
Db	58	SAGNTGNTGNPAKDNDNAGANDP	SKNDPSKDSQSPASANTGYVDNNDPDMQ	117
QY	43	LLAMFIMM	-----LQDSQSDANOEC	-----NEOPNQSQGLSPQLMQLMIVQMLQON- 94
Db	118	LLLEVLKILKALMHPQDNGDKNG	GVGGANGAGGCGGLGABAEQIILAJGGGG 177	
QY	95	-----	OGAGGCGGGSVNSIGN 113	
Db	178	AGAGGAGGGVGGAGDGGAGGAGG	GAGGAGDGN 212	
RESULT	6			
ID	QSWOP3	PRELIMINARY;	PRT; 318 AA.	
AC	QSWOP3			
DT	01-MAY-2000 (TEMBLrel. 13; Created)			
DE	01-MAY-2000 (TEMBLrel. 13; Last sequence update)			
DE	01-DEC-2000 (TEMBLrel. 19; Last annotation update)			
DE	CG17183 PROTEIN (SDS560P).			
GN	CG17183			
OS	Drosophila melanogaster (Fruit fly)			
OC	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;			
OC	Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;			
OC	Ephydroidae; Drosophilidae; Drosophila.			
OX	NCBI_TaxId=7227;			
RP	[1]			
RP	SEQUENCE FROM N.A.			
RC	SPRAIN-BERKELEY;			
RC	MDLIN20196006; PubMed=10731132;			
RA	Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,			
RA	Georgieva B., Giannakakis S., McEwen H.W., Hoskins R.A., Galle R.F.,			
RA	George R.A., Givens R.E., White R.L., Adams M.D., Chen L.,			
RA	Sutton G.G., Wortman J.R., Yeung M.P., Zeng K., Chen L., White O.,			
RA	Anderson R.C., Rogers Y.-H.C., Blazer V.G., Chang M., Pfeiffer B.D.,			
RA	Wan K.H., Doyle C., Baxter E.G., Holt G., Nelson C.R., Miklos G.L.G.,			
RA	Abri J.F., Agbayani A., An H.-J., Andrews-FannKoch C., Baldwin D.,			
RA	Bailly R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,			
RA	Beeson K.Y., Benos P.V., Berman B.P., Bhargava D., Bolshakov S.,			
RA	Borkova D., Botchan M.R., Bouck J., Brookstein P., Brotter P.,			
RA	Kurtz C.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,			
RA	Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,			
RA	deCaprio B.S., DeClerck A., Deng Z., Mays A.D., Dew I., Dietz S.M.,			
RA	Donnelly C., Doolittle R., Dugan-Rocha S., Dunkov B.C., Dunn P.,			
RA	Edwards J., Ewing R.C., Fagerberg S., Gage F., Gage R., Gieschmann W.,			
RA	Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasner K.,			
RA	Glisic A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., Houck J.,			
RA	Hodges N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,			
RA	Hoslin D., Houston K.A., Howland T.J., Wei M.-H., Ihebwan C.,			
RA	Jatelli M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,			
RA	Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,			
RA	Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,			
RA	Liu X., Mattai B., McIntosh T.C., McLeod M.P., McPherson D.,			
RA	Merkulov G., Milshina N.V., Mobarry C., Morris J., Moshrifi A.,			
RA	Mount S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,			
RA	Paletz M.M., Picman S., Pless S., Rasmussen R., Rhee M.G.,			
RA	Palazzo M., Picman S., Pless S., Rasmussen R., Rhee M.G.,			
RA	Shine B.C., Remington K., Saunders R.C., Schaefer P., Smith T.,			
RA	Rehner K., Siden-Klamis I., Simpson M., Skupski M.P., Smith T.,			
RA	Spter E., Spradling A.C., Stapleton M., Strong R., Sun E.,			
RA	Swirski R., Tector C., Turner R., Venter E., Wang A.H., Wang X.,			
RA	Wang Z.-Y., Wassman D.A., Weinstein G.M., Weissenbach J.,			
RA	Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,			
RA	Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,			
RA	Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu H.,			
RA	Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.			
PT	The genome sequence of <i>Drosophila melanogaster</i> ;			
PT	Science 287:2185-2195;(2000).			
RL	SEQUENCE FROM N.A.			
RA	Stapleton M., Brookstein P., Hong L., Agbayani A., Carlson J.,			
RA	Chamne W., Chavez-Dorsett Y., Fagerberg S., Gage R., Gieschmann W.,			

OY 61 NO-----ECNEQFON-----GQEGSLPTOMLQIVW-----Q 90
 DB 81 NEGQILGHVFGQOQAAANAIGRAGIDP--QMAQILMAVPLILSYLSNRSBAGGQ 138
 OY 91 LMONQGG-----AGMGG--CGSVNSLGG 112
 DB 139 MGNMGQSGGGMSLPCGMGGGLGGILGG 171

RESULT 10

OYV698 PRELIMINARY; PRT; 1366 AA.
 AC Q9V698
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DE MAN PROTEIN.
 GN MAN OR CG8118.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananthites P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Sutton G.G., Wortman J.R., Richards S., Ashburner M., Henderson S.N.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis J.M., Cawley S., Claessens D., Cledon E., Center A., Chandra I.,
 RA Cherry J.M., Chinai A., Dool L.E., Downes M., Dugan-Rocha S., Dunn P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Li J., Li Z., Liang Y., Lin X.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattal B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Pittman G.S., Pan S., Pollard J.D., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
 RA Shue B.C., Sidel-Kimos I., Simpson M., Skupski M.P., Strong R.,
 RA Spier E., Spradling A.C., Stapleton M., Turner E., Wang A.H., Wang X.,
 RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Zhis R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE001816; AAF58300.1;
 DR FlyBase: FBgn002643; man.
 SQ SEQUENCE 1366 AA; 141456 MW; CC10F8D5E23F3631 CRC64;

Query Match 16.5%; Score 96.5; DB 5; Length 1366;
 Best Local Similarity 28.8%; Pred. No. 0.073;

Matches 32; Conservative 7; Mismatches 39; Indels 33; Gaps 4;
 OY 22 QHEDSSQSFSGSAGSQQLOOLAMFPMHQLQSSDANQCGNPNQGGSLPT 81
 DB 1198 QQQQQQQHGFQFANMQLQLOO-----QSSGG-----GNMASQMTSH-MT 1244
 OY 82 QMLQIVMQLQIVW-----QSGMGGGGSVNSLGG 112
 DB 1245 GTQOOITMOOQOFGVOSTTTTTHQOOHMQHFGGSGGSGGSANNNGG 1295

RESULT 11

OYV697 PRELIMINARY; PRT; 1594 AA.
 AC Q9V697
 DT 01-MAY-2000 (TRENBLrel. 13, Created)
 DT 01-MAY-2000 (TRENBLrel. 13, Last sequence update)
 DE MAN PROTEIN.
 GN MAN OR CG8118.
 OS Drosophila melanogaster (fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BERKELEY.
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celniker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Ananthites P.G., Scher S.E., Li P.W., Hoskins R.A., Galle R.F.,
 RA Sutton G.G., Wortman J.R., Richards S., Ashburner M., Henderson S.N.,
 RA Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champagne M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA Abril J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Ballew R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Berman B.P., Bhandari D., Bolshakov S.,
 RA Borokova D., Botchan M.R., Bouck J., Brokstein P., Brotter P.,
 RA Burtis J.M., Cawley S., Claessens D., Cledon E., Center A., Chandra I.,
 RA Cherry J.M., Chinai A., Dool L.E., Downes M., Dugan-Rocha S., Dunn P.,
 RA de Pablo B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,
 RA Dodson K.J., Evangelista C.C., Ferraz C., Ferreira S., Fleischmann W.,
 RA Folsler C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K.,
 RA Glodek A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M.,
 RA Harris N.L., Harvey D., Heiman T.J., Hernandez J.R., Houck J.,
 RA Hostin D., Houston K.A., Howland T.J., Li J., Li Z., Liang Y., Lin X.,
 RA Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A.,
 RA Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z.,
 RA Lasko P., Lei Y., Levitsky A.A., Li J., Li Z., Liang Y., Lin X.,
 RA Liu X., Mattal B., McIntosh T.C., McLeod M.P., McPherson D.,
 RA Mout S.M., Moy M., Murphy B., Murphy L., Muzny D.M., Nelson D.L.,
 RA Nelson D.R., Pittman G.S., Pan S., Pollard J.D., Reese M.G.,
 RA Reinert K., Remington K., Saunders R.D.C., Scheeler F., Smith T.,
 RA Shue B.C., Sidel-Kimos I., Simpson M., Skupski M.P., Strong R.,
 RA Spier E., Spradling A.C., Stapleton M., Turner E., Wang A.H., Wang X.,
 RA Svirskas R., Tector C., Turner E., Venter E., Wang A.H., Wang X.,
 RA Williams S.M., Woodage T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao Q., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhou W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Zhis R.A., Myers E.W., Rubin G.M., Venter J.C.;
 RT "The genome sequence of Drosophila melanogaster."
 RL Science 287:2185-2195(2000).
 DR EMBL: AE001816; AAF58300.1;
 DR FlyBase: FBgn002643; man.
 SQ SEQUENCE 1594 AA; 167263 MW; D4236DA26F70D092 CRC64;

Query Match 16.5%; Score 96.5; DB 5; Length 1594;

us-09-829-124-2.rspt

DR EMBL: AF005221: AAC62526 1. -

Query Match 15.7% Score 91 5. DB 2: Testdb 124.

QY 26 DSSOQSPSAG-----SEOOOLDOLAMFTMMMI OOSGSGSDANOECCNEORONCAOECI SBT 90

RESULT 15

Q94486 ID Q94486 PRELIMINARY: PRT: 624 AA

Query Match

QY 22 00HEDSS00SPSAGSE00I.D0I.I.AMETM0MI 00S0CSDAN0ECGNE0P0NC00 74

Search completed: June 28, 2002 10:23:23

Job time: 150 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 28, 2002, 10:19:58 ; Search time 12.31 Seconds
(without alignments)

Title: US-09-829-124-2
 Perfect score: 584
 Sequence: 1 MDSIGNNFSNIGNLQTMGIG.....OGGAGMGGGGSVNSSILGNA 114

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	111	19.0	403	1	HRPN_ERWAM	Q01099 erwinia amy
2	110	18.8	344	1	POPA_RALSO	Q9rbs0 ralstonia s
3	96.5	16.5	1596	1	MAM_DROME	P21519 drosophila
4	89.5	15.3	597	1	IXR1_YEAST	P33417 saccharomyc
5	88	15.1	2414	1	P300_HUMAN	Q09472 homo sapien
6	87	14.9	727	1	BRC1_DROME	Q01295 drosophila
7	86.5	14.8	519	1	ELAV_DROVI	P23241 drosophila
8	85	14.6	675	1	ECR_AEADAE	P49880 aedes aegypt
9	83	14.2	353	1	TFD2_DROME	P20227 drosophila
10	82.5	14.1	1480	1	PAN1_YEAST	P32521 saccharomyc
11	81	13.9	660	1	GLT3_WHEAT	P08488 triticum ae
12	79	13.5	416	1	R23B_MOUSE	P54728 mus musculus
13	79	13.5	648	1	GLT0_WHEAT	P10387 triticum ae
14	78.5	13.4	512	1	FUS_BOVIN	Q28009 bos taurus
15	78	13.4	255	1	LP61_EIMTE	P15714 elmeria ten
16	78	13.4	1018	1	YRZ6_CAEEL	P34333 caenorhabdi
17	77	13.2	685	1	YICA_BACSU	P37483 bacillus su
18	76.5	13.1	405	1	YCC8_YEAST	P25367 saccharomyc
19	76.5	13.1	829	1	E74A_DROME	P20105 drosophila
20	76.5	13.1	883	1	E74B_DROME	P11536 drosophila
21	76	13.0	834	1	5HTA_DROME	P28285 drosophila
22	75	12.8	363	1	TOBI_MOUSE	Q61471 mus muscul
23	75	12.8	526	1	FUS_HUMAN	P35637 homo sapien
24	74.5	12.8	463	1	K10_DROME	P13468 drosophila
25	74.5	12.8	648	1	KAPC_DICDI	P34099 dictyosteli
26	74.5	12.8	1516	1	NC02_XENLA	Q9W705 xenopus lae
27	74.5	12.8	1845	1	R23B_HUMAN	Q9u136 homo sapien
28	74	12.7	409	1	R23B_HUMAN	P54727 homo sapien
29	73.5	12.6	738	1	YKF4_YEAST	P35732 saccharomyc
30	73.5	12.6	1043	1	FTFL_DROME	P33244 drosophila
31	73.5	12.6	2038	1	FSH_DROME	P13709 drosophila
32	73	12.5	269	1	BDNF_XIPMA	Q02193 xiphophorus
33	73	12.5	593	1	K1CJ_HUMAN	P13645 homo sapien

Query Match 19.0%; Score 111; DB 1; Length 403;

Query Match: 19.0%; Score III; DB I; Length 403;
Best Local Similarity 31.5%; Pred. No. 0.00071;

best local similarity 31.3%; Pred. No. 0.000/1;
Matches 39; Conservative 21; Mismatches 48; Indels

5 GUNFESNIGNI,-- --OTMGT-GPOOHEDSOOOOPS-ACSEOOIDOI AMETMMMO

```

||| :: | :|| |::| : | : ||| | : |
118 CANNYTCSTUPNCSPI DOATCINTECSANDBDCTSCMCMSCMCCSCDDMOOXY KMBCDTH

```

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59	60	61	62	63	64	65	66	67	68	69	70	71	72	73	74	75	76	77	78	79	80	81	82	83	84	85	86	87	88	89	90	91	92	93	94	95	96	97	98	99	100
0	0	1	4	9	16	25	36	49	64	81	100	121	144	169	196	225	256	289	324	361	400	441	484	529	576	625	676	729	784	841	900	961	1024	1089	1156	1225	1296	1369	1444	1521	1600	1681	1764	1849	1936	2025	2116	2209	2304	2401	2500	2601	2704	2809	2916	3025	3136	3249	3364	3481	3600	3721	3844	3969	4096	4225	4356	4489	4624	4761	4900	5041	5184	5329	5476	5625	5776	5929	6084	6241	6400	6561	6724	6889	7056	7225	7396	7569	7744	7921	8100	8281	8464	8649	8836	9025	9216	9409	9604	9801	10000

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840. 84

| | 73 | 12.5 | 905 | 1 | SNF5_YEAST | P18480 | saccharomyc |
|----|------|------|------|---|------------|--------|--------------|
| 34 | | | | | | | |
| 35 | 73 | 12.5 | 518 | 1 | FUS_MOUSE | P56959 | mus musculus |
| 36 | 72.5 | 12.4 | 700 | 1 | NONA_DROME | Q04047 | drosophila |
| 37 | 72.5 | 12.4 | 732 | 1 | KCN3_RAT | P70605 | rattus norv |
| 38 | 72 | 12.3 | 378 | 1 | HMAN_DROME | P28833 | drosophila |
| 39 | 72 | 12.3 | 401 | 1 | YK03_CAEEL | P34291 | caenorhabdi |
| 40 | 72 | 12.3 | 525 | 1 | NAB2_YEAST | P32505 | saccharomyc |
| 41 | 72 | 12.3 | 537 | 1 | YR32_EBV | P03210 | epstein-bar |
| 42 | 72 | 12.3 | 634 | 1 | HS70_CHICK | P08106 | gallus gall |
| 43 | 72 | 12.3 | 1160 | 1 | GLGI_CRIGR | Q9z1e9 | cricetulus |
| 44 | 72 | 12.3 | 1435 | 1 | NOS1_RABIT | O19132 | oryctolagus |
| 45 | 71.5 | 12.2 | 244 | 1 | GDB3_WHEAT | P04730 | triticum ae |

ALIGNMENTS

| RESULT | ID | HRPN_ERWAM | STANDARD; | PRT; | 403 AA. |
|--------|-------|---|-----------|------|---------|
| 1 | AC | Q01099; | | | |
| | DT | 01-APR-1993 (Rel. 25, Created) | | | |
| | DT | 30-MAY-2000 (Rel. 39, Last sequence update) | | | |
| | DT | 30-MAY-2000 (Rel. 39, Last annotation update) | | | |
| | DE | Harpin (Harpin-EA). | | | |
| | HRPN. | | | | |
| | OS | Erwinia amylovora. | | | |
| | OC | Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; | | | |
| | OC | Erwinia. | | | |
| | OX | NCBI_TaxID=552; | | | |
| | [1] | | | | |
| | KN | SEQUENCE FROM N.A., AND SEQUENCE OF 1-15. | | | |
| | RP | SEQUENCE-EA321; | | | |
| | RC | STRAIN-EA321; | | | |
| | RC | MEDLINE=92320301; PubMed=1621099; | | | |
| | RA | Wei Z.-M., Laby R.J., Zumoff C.H., Bauer D.W., He S.Y., Collmer A., | | | |
| | RA | Beer S.V.; | | | |
| | RT | "Harpin, elicitor of the hypersensitive response produced by the | | | |
| | RT | plant pathogen Erwinia amylovora." | | | |
| | RL | Science 257:85-88(1992). | | | |

12] REVISIONS.
RC STRAIN=EA321;
RA Laby R.J., Kim J.F., Beer S.V.;
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: ELICITS THE HYPERSENSITIVE RESPONSE (HR) IN THE PLANT
CC UPON INFECTION. HARPIN ELICITS HR IN NON-HOSTS AND IS ALSO
CC REQUIRED FOR PATHOGENICITY IN HOST PLANTS.
CC -!- SUBCELLULAR LOCATION: SECRETED; VIA THE HRP SECRETION PATHWAY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

| | | | | | |
|-----------------------------|---------|-----------|---------|-----------|--|
| EMBL; M92994; AAC31644.2; - | | | | | |
| Hypersensitive response. | | | | | |
| DOMAIN | 1 | 270 | | GLY-RICH. | |
| DOMAIN | 63 | 70 | | POLY-MET. | |
| SEQUENCE | 403 AA: | 39697 MW: | 146PA64 | | |

Query Match 19.0%; Score 111; DB 1; Length 403;

Query Match: 19.0%; Score III; DB I; Length 403;
Best Local Similarity 31.5%; Pred. No. 0.00071;

best local similarity 31.3%; Pred. No. 0.000/1;
Matches 39; Conservative 21; Mismatches 48; Indels

5 GUNFESNIGNI,-- --OTMGT-GPOOHEDSOOOOPS-ACSEOOIDOI AMETMMMO

```

||| :: | | :|| |::| : | : ||| | : |
118 CANNYTCSTUPNCSPI DOATCINTECSANDBDCTSCMCMSCMCCSCDDMOOXY KMBCDTH

```

[illegible]

1. 2. 3. 4. 5. 6. 7. 8. 9. 10. 11. 12. 13. 14. 15. 16. 17. 18. 19. 20. 21. 22. 23. 24. 25. 26. 27. 28. 29. 30. 31. 32. 33. 34. 35. 36. 37. 38. 39. 40. 41. 42. 43. 44. 45. 46. 47. 48. 49. 50. 51. 52. 53. 54. 55. 56. 57. 58. 59. 60. 61. 62. 63. 64. 65. 66. 67. 68. 69. 70. 71. 72. 73. 74. 75. 76. 77. 78. 79. 80. 81. 82. 83. 84. 85. 86. 87. 88. 89. 90. 91. 92. 93. 94. 95. 96. 97. 98. 99. 100. 101. 102. 103. 104. 105. 106. 107. 108. 109. 110. 111. 112. 113. 114. 115. 116. 117. 118. 119. 120. 121. 122. 123. 124. 125. 126. 127. 128. 129. 130. 131. 132. 133. 134. 135. 136. 137. 138. 139. 140. 141. 142. 143. 144. 145. 146. 147. 148. 149. 150. 151. 152. 153. 154. 155. 156. 157. 158. 159. 160. 161. 162. 163. 164. 165. 166. 167. 168. 169. 170. 171. 172. 173. 174. 175. 176. 177. 178. 179. 180. 181. 182. 183. 184. 185. 186. 187. 188. 189. 190. 191. 192. 193. 194. 195. 196. 197. 198. 199. 200. 201. 202. 203. 204. 205. 206. 207. 208. 209. 210. 211. 212. 213. 214. 215. 216. 217. 218. 219. 220. 221. 222. 223. 224. 225. 226. 227. 228. 229. 230. 231. 232. 233. 234. 235. 236. 237. 238. 239. 240. 241. 242. 243. 244. 245. 246. 247. 248. 249. 250. 251. 252. 253. 254. 255. 256. 257. 258. 259. 260. 261. 262. 263. 264. 265. 266. 267. 268. 269. 270. 271. 272. 273. 274. 275. 276. 277. 278. 279. 280. 281. 282. 283. 284. 285. 286. 287. 288. 289. 290. 291. 292. 293. 294. 295. 296. 297. 298. 299. 300. 301. 302. 303. 304. 305. 306. 307. 308. 309. 310. 311. 312. 313. 314. 315. 316. 317. 318. 319. 320. 321. 322. 323. 324. 325. 326. 327. 328. 329. 330. 331. 332. 333. 334. 335. 336. 337. 338. 339. 340. 341. 342. 343. 344. 345. 346. 347. 348. 349. 350. 351. 352. 353. 354. 355. 356. 357. 358. 359. 360. 361. 362. 363. 364. 365. 366. 367. 368. 369. 370. 371. 372. 373. 374. 375. 376. 377. 378. 379. 380. 381. 382. 383. 384. 385. 386. 387. 388. 389. 390. 391. 392. 393. 394. 395. 396. 397. 398. 399. 400. 401. 402. 403. 404. 405. 406. 407. 408. 409. 410. 411. 412. 413. 414. 415. 416. 417. 418. 419. 420. 421. 422. 423. 424. 425. 426. 427. 428. 429. 430. 431. 432. 433. 434. 435. 436. 437. 438. 439. 440. 441. 442. 443. 444. 445. 446. 447. 448. 449. 450. 451. 452. 453. 454. 455. 456. 457. 458. 459. 460. 461. 462. 463. 464. 465. 466. 467. 468. 469. 470. 471. 472. 473. 474. 475. 476. 477. 478. 479. 480. 481. 482. 483. 484. 485. 486. 487. 488. 489. 490. 491. 492. 493. 494. 495. 496. 497. 498. 499. 500. 501. 502. 503. 504. 505. 506. 507. 508. 509. 510. 511. 512. 513. 514. 515. 516. 517. 518. 519. 520. 521. 522. 523. 524. 525. 526. 527. 528. 529. 530. 531. 532. 533. 534. 535. 536. 537. 538. 539. 540. 541. 542. 543. 544. 545. 546. 547. 548. 549. 550. 551. 552. 553. 554. 555. 556. 557. 558. 559. 560. 561. 562. 563. 564. 565. 566. 567. 568. 569. 570. 571. 572. 573. 574. 575. 576. 577. 578. 579. 580. 581. 582. 583. 584. 585. 586. 587. 588. 589. 590. 591. 592. 593. 594. 595. 596. 597. 598. 599. 600. 601. 602. 603. 604. 605. 606. 607. 608. 609. 610. 611. 612. 613. 614. 615. 616. 617. 618. 619. 620. 621. 622. 623. 624. 625. 626. 627. 628. 629. 630. 631. 632. 633. 634. 635. 636. 637. 638. 639. 640. 641. 642. 643. 644. 645. 646. 647. 648. 649. 650. 651. 652. 653. 654. 655. 656. 657. 658. 659. 660. 661. 662. 663. 664. 665. 666. 667. 668. 669. 670. 671. 672. 673. 674. 675. 676. 677. 678. 679. 680. 681. 682. 683. 684. 685. 686. 687. 688. 689. 690. 691. 692. 693. 694. 695. 696. 697. 698. 699. 700. 701. 702. 703. 704. 705. 706. 707. 708. 709. 710. 711. 712. 713. 714. 715. 716. 717. 718. 719. 720. 721. 722. 723. 724. 725. 726. 727. 728. 729. 730. 731. 732. 733. 734. 735. 736. 737. 738. 739. 740. 741. 742. 743. 744. 745. 746. 747. 748. 749. 750. 751. 752. 753. 754. 755. 756. 757. 758. 759. 760. 761. 762. 763. 764. 765. 766. 767. 768. 769. 770. 771. 772. 773. 774. 775. 776. 777. 778. 779. 780. 781. 782. 783. 784. 785. 786. 787. 788. 789. 790. 791. 792. 793. 794. 795. 796. 797. 798. 799. 800. 801. 802. 803. 804. 805. 806. 807. 808. 809. 810. 811. 812. 813. 814. 815. 816. 817. 818. 819. 820. 821. 822. 823. 824. 825. 826. 827. 828. 829. 830. 831. 832. 833. 834. 835. 836. 837. 838. 839. 840. 84

Db 176 GQDGTGSSGKQTEGEQNAKKGVTDALSLMGNSQLLNGGLGGGGGAGTCL 235
QY 111 CGNA 114
Db 236 DGSS 239

RESULT 2
POPA_RALSO
ID POPA_RALSO STANDARD; PRT; 344 AA.
AC Q9RBS0; Q52544; Q9WKK0; Q9RBY0;
DT 01-MAR-2002 (Rel. 41, Created)
DT 01-MAR-2002 (Rel. 41, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE POPA1 protein [Contains: Popa2 protein; Popa3 protein].
GN POPA OR RSP0877 OR RS01648.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OG Plasmid megaplasmid.
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=94148001; PubMed=8313899;
RA Arlat M., Van Gijsegem F., Huet J.-C., Pernollet J.-C., Boucher C.A.;
RA "Popa1, a protein which induces a hypersensitivity-like response on
RT specific Petunia genotypes, is secreted via the Hrp pathway of
RT Pseudomonas solanacearum";
RL EMBO J. 13:543-553(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlat M., Billault A., Brottier P., Camus J.C., Cattolico L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunnac S., Demange N.,
RA Gaspin C., Lavie M., Moisan A., Robert C., Saurin W., Schiex T.,
RA Siguer P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RA "Genome sequence of the plant pathogen Ralstonia solanacearum";
RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=8107;
RA Hikichi Y., Ikegami M., Okuno T.;
RL Submitted (APR-1999) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=OEI-1;
RA Hikich Y., Kanda A., Hasegawa H., Okuno T.;
RT "POPA in Ralstonia solanacearum OEI-1 pathogenic to tobacco";
RL Submitted (SEP-1999) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: Act as a specific hypersensitive response (HR) elicitor.
CC Has activity on tobacco (non-host plant) and petunia but is
CC without activity on tomato (host plant).
CC -1- SUBCELLULAR LOCATION: Secreted.
CC -1- PTM: POPA2 and POPA3 are produced from popa1 by stepwise removal
CC of N-terminal amino acids.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; AJ245811; CAB58262.1;
CC EMBL; AL646081; CAD18028.1;
CC EMBL; AB026629; BAA77270.1;
CC EMBL; AB032747; BAA84679.1;
CC
CC Hypersensitive response; Plasmid; Complete proteome.

FT CHAIN 1 344 POPA1 PROTEIN.
FT CHAIN 10 344 POPA2 PROTEIN.
FT CHAIN 94 344 POPA3 PROTEIN.
FT VARIANT 24 26 TNS -> S (IN STRAIN OEI-1).
FT VARIANT 80 85 ANDPSK -> GA (IN STRAIN 8107).
FT VARIANT 173 173 L -> P (IN STRAIN OEI-1).
FT VARIANT 177 177 G -> A (IN STRAINS 8107 AND OEI-1).
FT VARIANT 187 196 MISSING (IN STRAIN 8107).
FT VARIANT 193 196 ADGG -> G (IN STRAIN OEI-1).
FT VARIANT 229 229 N -> I (IN STRAIN OEI-1).
FT VARIANT 269 269 G -> A (IN STRAIN 8107).
FT VARIANT 305 305 G -> D (IN STRAIN OEI-1).
FT VARIANT 329 329 A -> E (IN STRAIN OEI-1).
SQ SEQUENCE 344 AA; 33152 MW; 42A7C5A0C32B0907 CRC64;

Query Match 18.8%; Score 110; DB 1; Length 344;
Best Local Similarity 25.8%; Pred. No. 0.00075;
Matches 40; Conservative 16; Mismatches 55; Indels 44; Gaps 5;

QY 3 SIGNFNISNIGLQTM-----GIGPQHEDSSQSPSAGSE-----QQLDQ 42
Db 58 SAGGNTGTGNAPAKDGNANAGNDPDKNPSKSGQPSQANKTGNVDANNODPMQALMQ 117
QY 43 LLAMFIMM-----LQSSGSDANQECG--NEOPONGQEGSLPTLMQIMQIMQIMQ- 94
Db 118 LLEDLVLLKALHMQPGGNDKNGVGGANGAKGAGGGGLAEALQIEQLAQLGGGG 177
QY 95 -----QGGAGMGGGGVNSSLGNN 113
Db 178 AGAGGAGGVGGAGGADGGSGAGGANGADGNN 212

RESULT 3
MAM_DROME
ID MAM_DROME STANDARD; PRT; 1596 AA.
AC P21519;
DT 01-MAY-1991 (Rel. 18, Created)
DT 01-MAY-1991 (Rel. 18, Last sequence update)
DT 01-MAR-1992 (Rel. 21, Last annotation update)
DE Neurogenic protein mastermind.
GN MAM.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
OX NCBI_TaxID=7227;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CANTON-S;
RX MEDLINE=91065516; PubMed=1701150;
RA Smoller D., Friedel C., Schmid A., Bettler D., Lam L.,
RA Yedvobnick B.;
RT "The Drosophila neurogenic locus mastermind encodes a nuclear protein
RT unusually rich in amino acid homopolymers";
RL Genes Dev. 4:1688-1700(1990).
CC -1- FUNCTION: MAY HAVE A REGULATORY FUNCTION POSSIBLY IN ASSOCIATION
CC WITH THE N GENE PRODUCT.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- DEVELOPMENTAL STAGE: DURING EARLY NEUROGENESIS MAM PRODUCTS ARE
CC UBQUITOUSLY LOCATED. DURING LATER STAGES THEY ACCUMULATE IN THE
CC CENTRAL NERVOUS SYSTEM.
CC -1- MISCELLANEOUS: THE PROTEIN HAS MANY AA HOMOPOLYMERIC DOMAINS: 21
CC POLY-GLN RUNS (FROM 5 TO 16 AA IN LENGTH), 4 POLY-GLY (6 TO 10
CC AA), 3 POLY-ASN (3 X 5 AA), 1 POLY-ALA (10 AA) AND 1 POLY-THR (5
CC AA) RUNS.
CC -1- SIMILARITY: TO OTHER NUCLEAR PROTEINS OF DROSOPHILA, TO CERTAIN
CC YEAST AND MAMMALIAN REGULATORY PROTEINS.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way

| | | |
|------------|-------------|---------------------------------|
| RESULT | 5 | |
| P300_HUMAN | | |
| ID | P300_HUMAN | STANDARD; |
| AC | Q09472; | PRT; 2414 AA. |
| DT | 15-JUL-1998 | (Rel. 36, Created) |
| DD | 15-JUL-1998 | (Rel. 36, Last sequence update) |

| | | |
|------------|-------------|---------------------------------|
| RESULT | 5 | |
| P300_HUMAN | | |
| ID | P300_HUMAN | STANDARD; |
| AC | Q09472; | PRT; 2414 AA. |
| DT | 15-JUL-1998 | (Rel. 36, Created) |
| DD | 15-JUL-1998 | (Rel. 36, Last sequence update) |

| | | |
|------------|-------------|---------------------------------|
| RESULT | 5 | |
| P300_HUMAN | | |
| ID | P300_HUMAN | STANDARD; |
| AC | Q09472; | PRT; 2414 AA. |
| DT | 15-JUL-1998 | (Rel. 36, Created) |
| DD | 15-JUL-1998 | (Rel. 36, Last sequence update) |

01-MAR-2002 (Rel. 41, Last annotation update)
EIA-associated protein p300.
EF300 OR P300.
Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
NCBI_TaxID=9606;
[1]
SEQUENCE FROM N.A.
MEDLINE=95011587; PubMed=7523245;
RA Eckner R., Even M.E., Newsome D., Gerdes M., Decaprio J.A.,
RA Lawrence J.B., Livingston D.M.;
RT "Molecular cloning and functional analysis of the adenovirus EIA-
RT associated 300-kD protein (p300) reveals a protein with properties of
RT a transcriptional adaptor.";
RL Genes Dev. 8:869-884(1994).
CC -1- FUNCTION: PROBABLE TRANSCRIPTIONAL ADAPTOR REQUIRED FOR THE
CC ACTIVITY OF CERTAIN COMPLEX TRANSCRIPTIONAL REGULATORY ELEMENTS.
CC MAY HAVE A FUNCTION IN CELL CYCLE REGULATION. BINDS TO AND MAY BE
CC INVOLVED IN THE TRANSFORMING CAPACITY OF THE ADENOVIRUS EIA
CC PROTEIN.
CC SUBCELLULAR LOCATION: Nuclear.
CC -1- SIMILARITY: CONTAINS 1 BROMODOMAIN.
CC -1- SIMILARITY: CONTAINS 1 ZZ-TYPE ZINC FINGER.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and-for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).

EMBL; U01877; AAA18639.1; -
TRANSFAC; T01427; -
MIM; 602700; -
DR InterPro: IPR001487; Bromodomain.
DR InterPro: IPR003101; KIX.
DR InterPro: IPR000197; TAZ_finger.
DR InterPro: IPR000433; ZnF_ZZ.
DR Pfam; PF00439; bromodomain; 1.
DR Pfam; PF02172; KIX; 1.
DR Pfam; PF02135; zf-TAZ; 2.
DR Pfam; PF00569; ZZ; 1.
DR PRINTS; PR00503; BROMODOMAIN.
DR SMART; SM00297; BROMO; 1.
DR SMART; SM00291; ZnF_ZZ; 1.
DR PROSITE; PS00633; BROMODOMAIN_1; 1.
DR PROSITE; PS50014; BROMODOMAIN_2; 1.
DR PROSITE; PS01357; ZF_ZZ_1; 1.
DR PROSITE; PS0135; ZF_ZZ_2; 1.
KW Transcription regulation; Nuclear protein; Bromodomain; Cell cycle;
KW Zinc-finger.
KW KW
FT DOMAIN 11 17 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT FT FT
FT DOMAIN 1067 1139 BROMODOMAIN.
FT FT FT
FT DOMAIN 1572 1818 BINDING REGION FOR EIA ADENOVIRUS.
FT FT FT
FT ZN_FING 1664 1707 ZZ-TYPE.
FT FT FT
FT DOMAIN 797 800 POLY-SER.
FT FT FT
FT DOMAIN 1519 1526 POLY-GLU.
FT FT FT
FT DOMAIN 2066 2069 POLY-GLN.
FT FT FT
FT DOMAIN 2190 2195 POLY-GLN.
SQ SEQUENCE 2414 AA; 264143 MW; 6BFF909EE4B9D693 CRC64;

Query Match 15.1%; Score 88; DB 1; Length 2414;
Best Local Similarity 31.2%; Pred. No. 0.96;
Matches 39; Conservative 6; Mismatches 54; Indels 26; Gaps 5;

QY 13 NLATMGIG-----PQHEDSSQQSPSAGSEQQLDQ-----LLAMFIAMML 52
+ + + + + : : : : :
Db 2132 NMNPQAGVQRAGLPQQCPQQQLPPMGGSGSPAQQNMNHNTWTPSFQFRILLRQQHMQQ 2191
+ + + + + : : : : :
QY 53 QQSOGSDANDECG---NEQPONCQBEGLSLPTQHLMIQIVMLMONGGAGMGGSYNSS 109

Qy 65 GNEQPNGQEGLSPL- --TQMLMQIVQLMNQGAGMGGGGVNSSLGN 113
| : | | : | : | : | : | : | : | : | : | :
Db 564 GQQQPNIGILACSTPRKANTPTTQQQMYAAVMAAASASASTSGSANSSLNN 616

| RESULT | 7 | ELAV_DROVI | STANDARD; | PRT; | 519 AA. |
|--------|--|-----------------------------------|-----------|------|---------|
| ID | ELAV_DROVI | | | | |
| AC | P23241; | | | | |
| DT | 01-NOV-1991 | (Rel. 20, Created) | | | |
| DT | 01-NOV-1991 | (Rel. 20, Last sequence update) | | | |
| DT | 16-OCT-2001 | (Rel. 40, Last annotation update) | | | |
| DE | Elav protein (Embryonic lethal abnormal visual protein). | | | | |

OS *Drosophila virilis* (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 OX NCBI_TaxID=7244;

| | | |
|----|-----|-----------------------------------|
| RN | [1] | |
| RP | | SEQUENCE FROM N.A. |
| RX | | MEDLINE=91246165; PubMed=1903840; |

"Organizational analysis of elav gene and functional analysis of ELAV protein of Drosophila melanogaster and Drosophila virilis.";

CC - FUNCTION: PROBABLY INVOLVED IN THE RNA METABOLISM OF NEURONS.

CC
CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation,
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
CC or send an email to license@isb-sib.ch).

| | |
|----|--|
| DR | EMBL; M61748; AAA28505.1; . |
| DR | PIR; A40252; A40252. |
| DR | HSSP; P19339; 2SXL. |
| DR | FlyBase; FBgn0013110; Dvir\elav. |
| DR | InterPro; IPR002343; Hud_sxl_RNA. |
| DR | InterPro; IPR000504; RRM. |
| DR | Efam; PF00076; rtm; 3. |
| DR | PRINTS; PR00961; HUDSLRNA. |
| DR | SMART; SM00360; RRM; 3. |
| DR | PROSITE; PS50102; RRM; 3. |
| DR | PROSITE; PS00030; RRM_RNP_1; 2. |
| DR | RNA-binding; Repeat. |
| FT | DOMAIN 23 164 |
| FT | DOMAIN 185 276 |
| FT | DOMAIN 284 305 |
| FT | DOMAIN 438 516 |
| FT | SEQUENCE 519 AA; 55822 MW; B0IAF1C0D601B0D9 CRC64; |
| SQ | |

| | | | | |
|-----------------------|--------------|-----------------|----------------|--------------|
| Query Match | 14.8% | Score 86.5; | DB 1; | Length: 519; |
| Best Local Similarity | 27.8% | Pred. No. 0.25; | | |
| Matches 27; | Conservative | 14; | Mismatches 49; | Indels 7; |
| | | | | Gaps 2; |

Oy 14 LQTMGIGPQQHEDSSQQSPSAGSEQLDQLLAMFIMMLQQSQSDANQCENGPQNQG 73
 :| :|| : || :|| : | : || : || : || : || : || : || : || :
Dd .92 VQAAVVCCCCC0000000000VVQQGGGVVVQQAAVAVCC00000----000000000VVQ 146

QY 74 QEGLSPLTQMLMQIVMQLMQ--NQGAGMGGGGSVNS 108

Db 147 QQQQVVQQVQQVQQANTNGSGGAQNGSGSTET 183

| | |
|-----------|-----------|
| RESULT | 8 |
| ECR_AEDAE | |
| ID | ECR_AEDAE |
| STANDARD | |
| DDP | |
| SIZE | |

P49880;
01-OCT-1996 (Rel. 34, Created)
01-OCT-1996 (Rel. 34, Last sequence update)
15-JUL-1999 (Rel. 38, Last annotation update)
Ecdysone receptor (Ecdysteroid receptor) (20-hydroxy-ecdysone receptor) (20E receptor).
ECR OR NR1H1.
Aedes aegypti (Yellowfever mosquito).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Aedes.
OC NCBI_TaxID=7159;
OX [1]
RN SEQUENCE FROM N.A.
RP TISSUE=Fat body;
RC MEDLINE=95227190; PubMed=7711747;
RX "Cho W.L., Kapitskaya M.Z., Raikhel A.S.;
RT "Mosquito ecdysteroid receptor: analysis of the cdna and expression during vitellogenesis.";
RL Insect Biochem. Mol. Biol. 25:19-27(1995).
CC -1- FUNCTION: RECEPTOR FOR ECDYSONE. BINDS TO ECDYSONE RESPONSE ELEMENTS (ECRES).
CC -1- SUBUNIT: HETERODIMER OF USP AND ECR. ONLY THE HETERODIMER IS CAPABLE OF HIGH-AFFINITY BINDING TO ECDYSONE.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- TISSUE SPECIFICITY: FAT BODY AND OVARIES.
CC -1- DEVELOPMENTAL STAGE: A PEAK LEVEL EXPRESSION IS SEEN IN THE FAT BODY OF PREVITELLOGENIC FEMALE MOSQUITOS AT ONE AND TWO DAYS AFTER ECDYSTON. LEVELS FALL THREE-FOLD AT THREE DAYS POSTECLOSION. DURING THE VITELLOGENIC PERIOD, THERE IS A SMALLER PEAK IN THE FAT BODY AT 6 H PBM (6 HRS POST BLOOD MEAL). LATER LEVELS DECLINE AND REMAIN 2-2.5 FOLD LOWER THAN AT 6 H PBM THROUGHOUT THE REMAINDER OF THE VITELLOGENIC PERIOD UNTIL 48 H PBM WHEN IT IS NO LONGER DETECTABLE.
CC -1- SIMILARITY: BELONGS TO THE NUCLEAR HORMONE RECEPTORS FAMILY. NR1 SUBFAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

EMBL; U02021; AAA87394.1; --
HSPX; P20393; IA6Y.
InterPro: IPR000536; Hormone_rec_lig.
DR InterPro: IPR001723; Strdhormone_receptor.
DR InterPro: IPR001628; zf-C4.
DR Pfam; PF00104; hormone_rec; 1.
DR Pfam; PF00105; zf-C4; 1.
DR PRINTS; PR00398; STRDHORMONER.
DR PRINTS; PR00047; STROIDFINGER.
DR SMART; SM00430; HOLI; 1.
DR SMART; SM00399; ZNF_C4; 1.
DR PROSITE; PS00031; NUCLEAR_RECEPTOR; 1.
KW Receptor; transcription regulation; DNA-binding; Nuclear protein;
KW Zinc-finger.
KW DOMAIN 1 189 MODULATING (POTENTIAL).
FT DNA_BIND 190 255 NUCLEAR RECEPTOR-TYPE.
FT ZN_FING 190 210 C4-TYPE.
FT ZN_FING 226 250 C4-TYPE.
FT DOMAIN 348 569 HORMONE-BINDING (POTENTIAL).
FT DOMAIN 588 600 POLY-SER.
FT DOMAIN 616 648 GLN-RICH.
SQ SEQUENCE 675 AA; 74685 MW; 12986E8A76F4918 CRC64;

Query Match 14.6%; Score 85; DB 1; Length 675;
Best Local Similarity 29.6%; Pred. No. 0.47;
Matches 29: Conservative 10: Mismatches 39: Indels 20: Gaps 4:

| | | | |
|----|-----|---|----|
| OY | 22 | QHEDSSQSPASGSEQQQLDQLLAWFMIMMLQQSQG-SDANQECGNPONGQO-----E | 71 |
| Db | 580 | QMHSHTGTSIISSSSSSS-----SSNGSGNSGNSSSNQHPHPPHQ 626 | |
| OY | 76 | GLSLPTQMLMIVLMQMO-NOGGAGMGGGGVNSSLG 112 | |
| Db | 627 | QLTNPQOOHQOHQSOLQOVHANGSGGGGSNNSSGG 664 | |

| | | | |
|----------|---|-----------|--------------|
| RESULT 9 | | | |
| ID | TF2D_DROME | STANDARD; | PRT; 353 AA. |
| AC | P02027; Q9WZD8; | | |
| DT | 01-FEB-1991 (Rel. 17, Created) | | |
| DT | 01-FEB-1991 (Rel. 17, Last sequence update) | | |
| DE | Transcription Initiation factor TFIIID (TATA-box factor) (TATA | | |
| DE | sequence-binding protein) (TBP) (TATA-box binding protein). | | |
| DN | TBP OR TFIIID OR BTF1 OR CG9874. | | |
| OS | Drosophila melanogaster (Fruit fly). | | |
| OC | Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; | | |
| OC | Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; | | |
| OC | Ephydroidea; Drosophilidae; Drosophila. | | |
| OX | NCBI_TaxId=7227; | | |
| RP | [1] | | |
| RP | SEQUENCE FROM N.A. | | |
| RP | MEDLINE=90304877; PubMed=2194666; | | |
| RX | Hoey T., Dynlacht B.D., Peterson M.G., Pugh B.F., Tjian R.; | | |
| RT | "Isolation and characterization of the Drosophila gene encoding the | | |
| RT | TATA box binding protein, TFIIID."; | | |
| RL | Cell 61:1179-1186(1990). | | |
| RP | [2] | | |
| RP | SEQUENCE FROM N.A. | | |
| RP | MEDLINE=91067664; PubMed=2123550; | | |
| RX | Mulich M., Iida C.T., Horikoshi M., Roeder R.G., Parker C.S.; | | |
| RT | *cDNA clone encoding Drosophila transcription factor TFIIID.;" | | |
| RT | Proc. Natl. Acad. Sci. U.S.A. 87:9148-9152(1990). | | |
| RP | [3] | | |
| RP | SEQUENCE FROM N.A. | | |
| RP | MEDLINE=91067664; PubMed=2123550; | | |
| RX | Mulich M., Iida C.T., Horikoshi M., Roeder R.G., Parker C.S.; | | |
| RT | *cDNA clone encoding Drosophila transcription factor TFIIID.;" | | |
| RT | Proc. Natl. Acad. Sci. U.S.A. 87:9148-9152(1990). | | |
| RP | [3] | | |
| RP | SEQUENCE FROM N.A. | | |
| RP | Lira-Devito L.M., Burke T.W., Kadonaga J.T.; | | |
| RP | Submitted (JUN-1994) to the EMBL/GenBank/DBJ databases. | | |
| RP | [4] | | |
| RP | SEQUENCE FROM N.A. | | |
| RP | STRAIN=OREGON-R; | | |
| RC | Lee K., Oh Y., Yoon J., Cho N., Baek K.; | | |
| RA | Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases. | | |
| RP | [5] | | |
| RP | SEQUENCE FROM N.A. | | |
| RP | STRAIN=BERKELEY; | | |
| RC | MEDLINE=20196006; PubMed=10731132; | | |
| RA | Adams M.D., Celisner S.E., Holt R.A., Evans C.A., Gocayne J.D., | | |
| RA | Amanatides P.G., Scher S.E., Li P.W., Hoskins R.A., Galie R.F., | | |
| RA | George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N., | | |
| RA | Sutton G.C., Wortman J.R., Yeandle M.D., Zhang Q., Chen L.X., | | |
| RA | Brandon R.C., Rogers Y.-H.C., Blazej R.G., Champe M., Pfeiffer B.D., | | |
| RA | Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G., | | |
| RA | Abriel J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D., | | |
| RA | Balwle R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M., | | |
| RA | Beeson K.Y., Benos P.V., Beriman B.P., Bhandari D., Bolshakov S., | | |
| RA | Borkova D., Botchan M.R., Bouck J., Brockstein P., Brottier P., | | |
| RA | Burtis K.C., Busam D.A., Butler H., Cadieux E., Center A., Chandra I., | | |
| RA | Cherry J.M., Cavley S., Dahlke C., Davenport L.B., Davies P., | | |
| RA | de Pablo B., Delcher A., Deng Z., Mays-A.D., Dew I., Dietz S.M., | | |
| RA | Dodson K., Doup L.E., Downes M., Dugan-Rocha S., Dunkov B.C., Dunn P., | | |
| RA | Durbin K.J., Evangelista C.C., Ferraz C., Ferrieria S., Fleischmann W., | | |
| RA | Foster C., Gabrielian A.E., Garg N.S., Gelbart W.M., Glasser K., | | |
| RA | Glocke A., Gong F., Gorrell J.H., Gu Z., Guan P., Harris M., | | |
| RA | Harris N.L., Harvey D., Helman T.J., Hernandez J.R., Houck J., | | |
| RA | Hostin D., Houston K.A., Howland T.J., Wei M.-H., Ibegwam C., | | |
| RA | Jalali M., Kalush F., Karpen G.H., Ke Z., Kennison J.A., Ketchum K.A., | | |
| RA | Kimmel B.E., Kodira C.D., Kraft C., Kravitz S., Kulp D., Lai Z., | | |
| RA | Lasko P., Lei Y., Levitsky A., Li J., Li Z., Liang Y., Lin X., | | |
| RA | Liu X., Mattel B., McIntosh T.C., McLeod M.P., McPherson D., | | |

RT HMW-glutenin subunit.
 CC Nucleic Acids Res. 13:6833-6846(1985).
 CC -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
 CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE
 CC VISCO-ELASTIC PROPERTY OF WHEAT DOUGH.
 CC -!- SUBUNIT: DISULFIDE-BRIDGE LINKED AGGREGATES.
 CC -!- MISCELLANEOUS: GLUTENINS ARE CODED BY SEVERAL GENES ON EACH OF THE
 CC GROUP 1 CHROMOSOMES OF WHEAT.
 CC -!- MISCELLANEOUS: THE MATURE PROTEIN IS CHARACTERIZED BY A LARGE
 CC NUMBER OF WELL PRESERVED REPEATS OF THE TWO MOTIFS: GQDPGQ AND
 CC GQDPGQGGQGYPTS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X03041; CAA26847.1;
 DR PIR; A24266; A24266.
 DR HSSP; P01088; 1BFA.
 DR InterPro; IPR001768; Cereal_tryp_amyl_inh.
 DR InterPro; IPR001419; Glutenin.
 DR Pfam; PF00234; tryp-alpha_amyl; 1.
 DR PRINTS; PR00210; GLUTENIN.
 KW Seed storage protein; Repeat; Multigene family; Signal.
 FT SIGNAL 1 21
 FT CHAIN 22 660
 FT DOMAIN 125 615
 FT REPEATS.
 FT SEQUENCE 660 AA; 70868 MW; 2BFD09D8C8FCCCF CRC64;

Query Match 13.9%; Score 81; DB 1; Length 660;
 Best Local Similarity 27.4%; Pred. No. 1.1;
 Matches 29; Conservative 10; Mismatches 45; Indels 22; Gaps 4;
 QY 12 GNLQTMIGPQHQHEDSSQSPSAGSEQQLDQLLAMFTMMMLQSQGSDANQECNEQPN 71
 Db 264 GQWQSGQGGQGHYPTSLQPGGQ-----GHYLSAQQPAQGGQGHYPASQGGQPGQ 316
 QY 72 GQDEGLSPLTQMLQIVMLQMQGG-----AGMGSGGSVNS 109
 Db 317 GQQ-GHYPASQ-----QDPGQGGHYPASQGPQGGQGGQIPAS 355

RESULT 12
 ID R23B_MOUSE STANDARD; PRT; 416 AA.
 AC P54728;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE UV excision repair protein RAD23 homolog B (MHR23B) (XP-C repair
 DE complementing complex 58 kDa protein) (P58).
 GN RAD23B OR MHR23B.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=BAUB/C; TISSUE=Testis;
 RX MEDLINE=96403997; PubMed=8808275;
 RA van der Spek P.J., Visser C.E., Hanaoka F., Smit B.,
 RA Hagemeijer A., Bootsma D., Hoeijmakers J.H.J.;
 RT "Cloning, comparative mapping, and RNA expression of the mouse
 RT homologues of the Saccharomyces cerevisiae nucleotide excision repair
 RT gene RAD23.";
 RL Genomics 31:20-27(1996).
 CC -!- FUNCTION: INVOLVED IN DNA EXCISION REPAIR. MAY PLAY A PART IN DNA

CC DAMAGE RECOGNITION AND/OR IN ALTERING CHROMATIN STRUCTURE TO
 CC ALLOW ACCESS BY DAMAGE-PROCESSING ENZYMES.
 CC -!- SUBUNIT: HETERODIMER OF A 125 kDa SUBUNIT (P125) AND OF A
 CC 58 kDa SUBUNIT (P58).
 CC -!- SUBCELLULAR LOCATION: Nuclear (Probable).
 CC -!- SIMILARITY: CONTAINS 1 UBIQUITIN-LIKE DOMAIN.
 CC -!- SIMILARITY: CONTAINS 2 UBA DOMAINS.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X92411; CAA63146.1;
 DR HSSP; P54725; 1DV0.
 DR MGI; MGI:105128; Rad23b.
 DR InterPro; IPR000449; UBA.
 DR InterPro; IPR000626; Ubiquitin.
 DR Pfam; PF00627; UBA; 2.
 DR Pfam; PF00240; ubiquitin; 1.
 DR SMART; SM00165; UBA; 2.
 DR SMART; SM00213; UBO; 1.
 DR PROSITE; PS00553; UBIQUITIN_2; 1.
 KW DNA damage; DNA repair; Nuclear protein.
 FT DOMAIN 1 79 UBIQUITIN-LIKE.
 FT DOMAIN 255 261 POLY-ALA.
 FT DOMAIN 262 270 POLY-THR.
 FT DOMAIN 336 355 POLY-GLY.
 FT SEQUENCE 416 AA; 43516 MW; 13E0245A6D892205 CRC64;

Query Match 13.5%; Score 79; DB 1; Length 416;
 Best Local Similarity 30.8%; Pred. No. 1.1;
 Matches 28; Conservative 9; Mismatches 30; Indels 24; Gaps 3;
 QY 22 QQHEDSSQSPSAGSEQQLDQLLAMFTMMMLQSQGSDANQECNEQPNQGGGLSPLT 81
 Db 286 QQMROITIQNPSTL-----LPALL-----QQIGRENPLQLQ--ISQHQ 321
 QY 82 QMLMQLVQMLQMONQGGAGMGSGSVNSLGG 112
 Db 322 EHFQMLNEPQVEAGGGGGGGGGGGGGG 352

RESULT 13
 ID GLT0_WHEAT STANDARD; PRT; 648 AA.
 AC P10387;
 DT 01-MAR-1989 (Rel. 10, Created)
 DT 01-MAR-1989 (Rel. 10, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE Glutenin, high molecular weight subunit D110 precursor.
 GN GLU-D1-2B.
 OS Triticum aestivum (Wheat).
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Pooidae;
 OC Triticeae; Triticum.
 OX NCBI_TaxID=4565;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=CV. CHEYENNE;
 RX MEDLINE=89098419; PubMed=2563152;
 RA Anderson O.D., Greene F.C., Yip R.E., Halford N.G., Shewry P.R.,
 RA Malpica-Romero J.M.;
 RT "Nucleotide sequences of the two high-molecular-weight glutenin genes
 RT from the D-genome of a hexaploid bread wheat, Triticum aestivum L. cv
 RT Cheyenne.";
 RL Nucleic Acids Res. 17:461-462(1989).
 CC -!- FUNCTION: GLUTENINS ARE THE HIGH MOLECULAR WEIGHT SEED STORAGE
 CC PROTEIN OF WHEAT ENDOSPERM. THOUGHT TO BE RESPONSIBLE FOR THE

RC TISSUE-Sporozoite;
RX MEDLINE-30348718; PubMed-2200963;
RA KO C., Smith C.K. II, McDonnell M.;
RT "Identification and characterization of a target antigen of a
monoclonal antibody directed against *Eimeria tenella* merozoites.";
RL Mol. Biochem. Parasitol. 41:53-64(1990).
CC -1- FUNCTION: UNKNOWN. THE GLN-RICH TANDEM REPEATS MAY BE IMPORTANT
FOR AN UNKNOWN ASPECT OF THE PARASITIC LIFE CYCLE. MAY BE AN
CC IMPORTANT IMMUNOGEN.
CC -1- SUBUNIT: MAY BE COVALENTLY LINKED BY DISULFIDE BONDS TO OTHER
CC POLYPEPTIDES TO FORM THE 80 KDa ANTIGEN.
CC -1- DEVELOPMENTAL STAGE: PRESENT IN ALL STAGES THROUGHOUT THE
CC SPORULATION OF THE OOCYSTS AND IN THE SPOOROZOITES FOLLOWING
CC EXCYSTATION.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M30933; AAA29079.1;
DR PIR; A60637; A60637.
KW Antigen; Sporozoite; Repeat; Sporulation.
FT NON_TER 1
FT DOMAIN 18 210 12 X APPROXIMATE TANDEM REPEATS, GLN-
RICH.
FT REPEAT 18 48 1.
FT REPEAT 49 57 2.
FT REPEAT 58 65 3.
FT REPEAT 66 78 4.
FT REPEAT 79 90 5.
FT REPEAT 91 103 6.
FT REPEAT 104 140 7.
FT REPEAT 141 152 8.
FT REPEAT 153 164 9.
FT REPEAT 165 172 10.
FT REPEAT 173 192 11.
FT REPEAT 193 210 12.
FT NON_TER 255 255
SQ SEQUENCE 255 AA; 31267 MW; 8C5E6005FFFC2DB3 CRC64;

Query Match 13.4%; Score 78; DB 1; Length 255;
Best Local Similarity 32.6%; Pred. No. 0.79;
Matches 30; Conservative 6; Mismatches 48; Indels 8; Gaps 2;
QY 22 QQHEDSSQSPSAGSEQQDQLLAFIMMLOQSGSDANECGNPQNGQEGLSPLT 81
DB 144 000000000WPEQEQEQEQEQWPEQ-----QQQQWSDNQQQQAGQWQAQQQQWPPQQ 197
QY 82 QMLMQIVMLQMGAGMGCGGSGVNSLGGN 113
DB 198 QQPQQ--QQQQQQQDPLGPDGVGVVPLGSS 227

Search completed: June 28, 2002, 10:22:46
Job time: 168 sec